



Using Machine Learning to Predict Port Congestion

A study of the port of Paranaguá

Eirik Fadnes & Espen Åsheim Harviken

Supervisors: Haiying Jia & Roar Os Ådland

Master thesis, Economics and Business Administration

Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

Acknowledgements

We would like to thank our supervisors, Haiying Jia and Roar Os Ådland, for sharing their extensive knowledge about the shipping industry with us. Furthermore, their guidance and advice throughout this process has been very valuable. Secondly, we want to show our utmost appreciation to Phil Curran and Trond Aga Haug at G2 Ocean for valuable insight into their operations and for granting us the opportunity to write our thesis in collaboration with their company. The authors thank Gabriel M. Fuentes for sharing data with us, as well as the code used for his Ph.D. dissertation. Lastly, we offer our thanks to K7 Minutter for letting us use their office and workstation.

Norwegian School of Economics

Bergen, June 2023

Eirik Fadnes

Espen Åsheim Harviken

Abstract

Being able to accurately predict future levels of port congestion is of great value to both port and ship operators. However, such a prediction tool is currently not available. In this thesis, a Long Short-Term Memory Recurrent Neural Network is built to fulfill this need. The prediction model uses information mined from Automatic Identification Systems (AIS) data, vessel characteristics, weather data, and commodity price data as input variables to predict the future level of congestion in the port of Paranaguá, Brazil. All data used in this study are publicly available. The predictions of the proposed model are shown to be promising with a satisfactory level of accuracy. The conclusion and evaluation of the presented model are that it serves its purpose and fulfills its objective within the constraints set by the authors and its inherent limitations.

Keywords – Shipping Industry, Port Congestion, Machine Learning, Long Short-Term Memory Recurrent Neural Network, Weather Margin, Dijkstra’s Algorithm

Contents

1	Introduction	1
2	Background	3
3	Literature Review	5
3.1	Port Operations and Congestion	5
3.2	Weather Effect	7
3.3	Machine Learning	8
3.4	Performance Metrics	9
4	Data	11
4.1	AIS Data	11
4.2	Meteorological Data	12
4.3	Additional Data	13
4.4	Data Preprocessing	14
4.4.1	Processing AIS Data	14
4.4.2	Weather Data Encoding	17
5	Methodology	18
5.1	Spatiotemporal Data Mining	18
5.1.1	Vessel Turnaround Time	18
5.1.2	Synthesizing ETA with Dijkstra’s Algorithm	20
5.2	LSTM Neural Network	21
5.3	Multi-Step Multivariate LSTM Model	23
5.3.1	Data Preparation for LSTM	24
5.3.2	Model Training	26
5.3.3	Hyperparameter Optimization	26
5.4	Model Structure	27
6	Analysis and Discussion	29
6.1	Model Predictions	29
6.2	Performance Evaluation	30
6.2.1	Comparing with Benchmark Models	31
6.2.2	Feature Importance	33
6.3	Operational Implications	35
7	Limitations and Further Research	37
7.1	Limitations	37
7.2	Further Research	38
8	Conclusion	39
	References	40
	Appendix	46
A1	Methodology Illustration	46

List of Figures

2.1	A histogram of turnaround time for bulk carriers visiting Paranaguá. . .	3
4.1	Distance (nautical miles) and estimated TTA using reported AIS ETA. TTA is calculated as the difference between the start of a trip and reported ETA. Note that all negative values are faulty.	16
5.1	A map of the Paranaguá port area with the geofenced areas of interest; the berthing zone where vessels are in port, the anchoring zone where vessels are waiting to enter the port, and the navigation zone where vessels are moving between them. Also on the map is the port of São Francisco do Sul with corresponding anchoring cluster, south of Paranaguá.	19
5.2	An STL-decomposition of the last three years showing the linear components of turnaround. The full turnaround can be seen in the top panel.	20
5.3	A map of ships traveling to Paranaguá color-coded to show estimated time-to-arrival.	21
5.4	A diagram showing the structure of a long short-term memory neural network.	22
5.5	Sliding window technique for one week prediction. An input sequence of 600 data points is used to predict the output sequence of 300 points, before sliding one data point and repeating the process.	25
6.1	Predictions using LSTM with different forecast horizons on the test data. The length of the test set varies slightly due to the differing lengths of the input for each prediction point.	29
6.2	sMAPE for different forecasting horizons, one to eight weeks.	31
6.3	sMAPE for the XGBoost, Naïve model and LSTM with different forecasting horizons.	32
6.4	A beeswarm plot of SHAP values calculated for each variable using the SHAP library developed by Lundberg and Lee (2017).	34
6.5	MAE for different forecasting horizons, one week to eight months. The MAE score is interpreted as the average error for predicted turnaround time in hours.	36
A1.1	Methodology inspired by Abebe et al. (2020)	46

List of Tables

4.1	AIS Data Fields	12
5.1	Complete Dataset	25
5.2	Structure of LSTM Models	28
6.1	LSTM Models Performances	30
6.2	Model Performance Comparison with sMAPE	32

1 Introduction

Ports are an integral part of global maritime shipping. They play a critical role in supporting global supply chains by ensuring that the inflow and outflow of ships go as efficiently as possible, which is not an easy task. According to UNCTAD (2019), the global median time ships spent in ports in 2018 was 23.5 hours. As a consequence of covid-19 and following increased pressure on many supply chains, the median port time increased through 2019 and 2020 and was measured in 2021 at 25.2 hours (Sirimanne et al., 2019; Notteboom et al., 2021; Gui et al., 2022). Increased time spent in ports causes congestion, which can cause significant delays, increased fuel consumption, and even accidents for ship operators (Meersman et al., 2012). Furthermore, 93.6% of the delays that occur in ports are due to congestion (Notteboom, 2006). Therefore, being able to accurately predict future levels of congestion will be largely beneficial for both port operators and shipping companies during normal operating conditions and also when unexpected events occur.

G2 Ocean, a joint venture company between Gearbulk and Grieg Maritime Group (G2 Ocean AS, 2023), has long experience with port congestion and is looking to improve its ability to predict future congestion for its fleet. Predicting future port congestion will help the company limit costs and reduce risk by adjusting speed and negotiating better terms in their contracts. G2 Ocean has become the world's largest ship operator in the open hatch segment, covering 37 trade lanes around the world (G2 Ocean AS, 2022). In 2021, their ships were present in ports more than 3,700 times in more than 60 countries, making it necessary to know the situation at the next port.

There are many sources of congestion in ports. Uncertainties with port productivity will always be present and can lead to substantial delays in operations. Congestion can occur because ships have to wait until a berth becomes available, other ships occupy loading-discharge facilities, or the entrance or departure gates of the port are congested (Pruyn et al., 2020). Such problems occur frequently, and previous studies have suggested probability modeling to combat this measurable uncertainty (Notteboom, 2006). However, there exist variables affecting port congestion that are not related to port operations themselves. Weather at sea affects how quickly ships travel and what route will be optimal to take (Park et al., 2021; Zis et al., 2020). The increased frequency and relative force of

extreme weather (Portner et al., 2019), make weather variables a greater potential source of congestion and therefore must be included when predicting future congestion.

Little previous research has been dedicated to predicting port congestion. Most studies have focused only on measuring congestion, in part due to poor access to high-quality dynamic data (Yeo et al., 2007; Steven and Corsi, 2012; Leachman and Jula, 2011; Oyatoye et al., 2011). This data access problem can now be mitigated by using data from the Automatic Identification System (AIS). AIS is a vessel tracking system and anticollision system that automatically tracks and shares updates on a vessel’s position and other voyage information. Ships equipped with an AIS transponder transmit a packet every few seconds, containing dynamic data such as the vessel’s position, speed over ground, course over ground, true heading, and rate of turn. Static vessel-specific data, draft, and other information related to the voyage are transmitted less frequently.

The work of Peng et al. (2022) and Abualhaol et al. (2018a) is used as a basis for the methodology of this thesis, with regard to the use of deep learning and long short-term (LSTM) neural network for actual predictions. Furthermore, the thesis takes inspiration from the work of Fuentes (2021), who clustered vessels according to their position in and around the port area to detect whether the vessel is in berth, anchorage, or moving.

In addition, this thesis contributes to the existing literature by paying more attention to the importance of weather variables and other variables with a potential causal effect on the formation of port congestion. In addition to AIS data, we also collected weather data from the Copernicus Climate Change Service database, commodity prices from NASDAQ and Bloomberg, and ship characteristics from Clarksons’ World Fleet Register. The goal of this thesis is to develop a model capable of predicting future port congestion using large amounts of publicly available data. To the best of our knowledge, this paper marks the first application of LSTM RNNs for long-term prediction of port congestion.

In agreement with G2 Ocean, we limit the scope of the thesis to predict congestion in a single port, Paranaguá port in Brazil. Paranaguá is the second largest port in Brazil and is frequently visited by the G2 Ocean fleet. The limited scope should not be seen as a limitation of the model output. Rather, it should be seen as a proof of concept that such a prediction model is possible to construct and will produce valuable information for stakeholders in the shipping industry.

2 Background

As G2 Ocean is one of the world’s largest shipping companies with operations spanning across multiple regions, various unforeseen disruptions are likely to occur. The entire maritime shipping industry experienced this in 2021 when the Ever Given vessel, operated by the company Evergreen, had an accident and blocked the Suez Canal. The canal is one of the most used shipping routes, where 13.3% of the world’s freight is transported through it, making it a crucial part of many supply chains (Lee and Wong, 2021). This led to an immense propagation effect among the following vessels and congestion in ports. In that situation, a tool with real-time data and forecasts of congestion would be of great value to all incoming vessel operators.

As mentioned, we limit our study to the port of Paranaguá in Brazil. It is located in the Paranaguá bay, just south of São Paulo, which is the largest exporter of agricultural products in the country and the largest bulk port in Latin America (World Port Source, 2023). Additionally, the layout of the port makes it so that the awaiting vessels are anchored just outside of the bay. The port has 16 berths for commercial vessels (container and bulk), four berths for liquid cargo and two berths for fertilizer (Portos do Paraná, Port Authority, 2023).

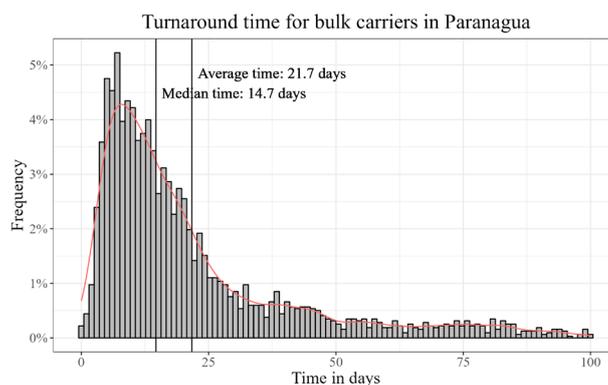


Figure 2.1: A histogram of turnaround time for bulk carriers visiting Paranaguá.

G2 Ocean has a long history with the port of Paranaguá and has served several local companies with frequent sailings, mostly transporting goods to the United States and northern Europe. The main commodities loaded onto the G2 Ocean vessels in Paranaguá are wood pulp, kraft liner board, and construction lumber. Currently, the demand is three shipments per month and approximately 50 megatons (Mt) are being shipped by

G2 Ocean in the Paranaguá port. The company not only ships goods from Brazil but also brings cargo to the country. However, this occurs mainly during certain seasonal periods and the cargo is usually fertilizer needed for agricultural production in Brazil. It is especially during these periods of elevated demand that the company experiences an increase in port congestion in Paranaguá, which is due to the high increase in fertilizer transportation to the port.

During the peak fertilizer season, G2 Ocean often evaluates the potential for congestion as too risky if they are booked on liner terms. Meaning they operate on a schedule with fixed port rotation at a fixed frequency (Shipping and Freight Resouce, 2023). In these cases, the company looks to nearby ports such as Itajaí port and Sao Francisco do Sul port to redirect its cargo. Therefore, it is critical that G2 Ocean is prepared for future congestion to negotiate terms and space allocation in alternative ports.

To ensure that port congestion predictions are of best use, G2 Ocean has requested a time span of two to three months for the prediction. Being able to predict future congestion two to three months in advance will allow the company to delay the sailing of vessels in case of unusually long waiting time at the port of destination. However, predictions within a shorter time frame, when the vessel is already at sea, will still be useful since the ship operators can adjust the speed of the vessel to sail most cost-efficiently. G2 Ocean recognizes the potential to exploit big data and acknowledges the availability of information. However, until recently, the resources and methods necessary to extract meaningful patterns have been lacking.

The following sections of the thesis are structured as follows. Section 3 reviews the literature on port congestion measurement and prediction, the effect of weather on shipping, and the use of machine learning in this field. The following section provides an explanation of the different data used in this thesis and the preprocessing that is performed to utilize the data. Section 5 provides an explanation of the methodological framework and approach used in this study, followed by an analysis of the results of the prediction model developed. Afterwards, we discuss the findings in the Discussion section and disclose the limitations with our approach. Finally, we present our conclusions in Section 8.

3 Literature Review

3.1 Port Operations and Congestion

The transportation of cargo from an origin country to a destination country involves three key components: the movement of the cargo to and from the outbound and inbound sea ports, the handling and loading onto the ship within the port terminals, and the oceanic journey of the cargo onboard a vessel. At the end of the shipment, the cargo is unloaded from the ship and trans-loaded onto other means of transport. When this process is seen as a system, it becomes clear that the sea port plays a crucial role at both the origin and destination ends. Inefficiencies in ports, such as slow loading/unloading, workforce problems, or limited connections to land transportation, significantly detract from the overall efficiency of the transportation process. In addition, ships arriving at the port at other times than planned also contribute to increased uncertainty and inefficiencies. The direct consequence of all this is, ultimately, queueing, longer waiting times, and congestion (Meersman et al., 2012).

Other researchers have used different ways to measure and quantify port congestion. Yeo et al. (2007) estimated and simulated traffic congestion in the Busan port by calculating the number of vessels waiting, the time interval between the arrival of the vessel and the time of service on the berth. Another attempt to measure congestion was made by Steven and Corsi (2012) who used a relationship between the total number of container vessel calls and the total number of container berths available in the port. However, this measure will only cover the congestion present at the berth, so they add a second measure using the ratio of number of containers that are lifted over number of cranes at the port. Leachman and Jula (2011) calculated container flow times in US ports and used them as a measure of congestion where ports with higher flow times were more congested. All of these studies have used data from port authorities or industry players, making the data heterogeneous and prone to inconsistencies. Furthermore, the articles maintain a limited view of congestion and do not account for other plausible sources of congestion.

With access to more accurate, homogeneous and consistent data along with the emergence of machine learning, researchers have been able to measure, better categorize, and predict

future port congestion. Abualhaol et al. (2018a) utilizes AIS data to accurately measure congestion pressure in different parts of port areas. They developed three port congestion indicators that provide an accurate representation of congestion. The indicators are spatial complexity, spatial density, and time criticality. Although the researchers did not predict future congestion in this article, they did present a set of indicators that can be used for predictions. Note that the authors did not include any other data in their model than those that can be found in the AIS data. We aim to include other data on port congestion in this thesis.

In a later paper, Abualhaol et al. (2018b) used their set of congestion indicators to forecast future port congestion in the port of Singapore. They do this using an LSTM Recurrent Neural Network (RNN) model and their prediction results were promising, demonstrating that such a prediction model using AIS data can be used in congestion prediction with confidence.

The precision of AIS data has made it possible to precisely detect different stages of port operations, from berthing, anchorage, and waiting to vessels passing by. Fuentes (2021) developed a method to dynamically identify ships anchored within a port area, as well as being served at the berth area using AIS data. Fuentes makes use of the density-based spatial clustering of applications with noise (DBSCAN) algorithm. The algorithm classifies positions (points) into clusters based on parameters selected by the user (Ester et al., 1996).

Peng et al. (2022) continued on the path started by Abualhaol et al. (2018a) and Fuentes (2021). They also used AIS data and created clusters of berths and anchorages in 20 of the largest ports in the world. To measure congestion, they used two dimensions: traffic volume and turnaround time. Contrary to Fuentes (2021), Peng et al. built a prediction model for actual hourly congestion and used an LSTM RNN to do so. Like the work of Abualhaol et al. (2018b), the results of their predictions are promising. However, they find that individual port congestion measures perform best when developed with respect to port-specific attributes.

In their study of 2022, Bai et al. clustered and located individual ports without differentiating between berth and anchorage. The study did not aimed to predict congestion, but only to measure it and investigate its economic impact. To quantify and

measure port congestion, the authors calculated the time span from the vessel's entry into the recognized port area to its exit. Subsequently, they defined actual congestion by the difference between the actual observed time in the port and the normally required time for loading and discharging, which they had access to. The normally required time for loading and discharging is data we do not have access to for the vessels in our dataset. Our data are relatively heterogeneous consisting of vessels of different sizes and storage capabilities, and we do not know the volume of loading/unloading per ship. However, this is desired to model reality with a high degree of accuracy.

3.2 Weather Effect

The effect of different weather conditions on the speed of a ship, the chosen route, and ultimately the travel time is well documented (Kuroda and Sugimoto, 2022; Adland et al., 2015; Zis et al., 2020; Taskar and Andersen, 2020; Nilsson and Nilsson, 2021; Knive and Liane, 2022). Ships at sea facing severe weather conditions must adjust their movements accordingly to reduce risk (Adland et al., 2015; Heij and Knapp, 2015). Adjustments to speed and course will bring additional uncertainty to the estimated arrival time (ETA) of the vessel at its destination and may result in operational delays at the port. Furthermore, since each ship has its own route, they will experience different external operating conditions in space and time (Ando, 2014). Therefore, we see meteorological data as a critical factor to include in a prediction model on future port congestion.

The use of meteorological and oceanographic data in ship routing has received increasing attention in recent years. Zis et al. (2020) conducted a meta-study of research articles focused on the area of weather routing and voyage optimization. They found that weather routing has evolved over the years, and earlier strategies have mainly focused on avoiding areas with severe storms. However, more complex objectives have emerged as technology and data quality have improved. Today, weather forecasting provides more detailed and accurate information with better spatial and temporal resolution. This improved data quality and availability can aid in the development of algorithms to determine the optimal path for a voyage. Additionally, the widespread use of AIS generates vast amounts of data on ships' sailing speed and position. This can be used to track deviations from traditional routes due to weather and measure the actual benefits of weather routing, particularly in

terms of saving total voyage time.

Mao et al. (2016) used weather and oceanographic information along with data on the engine performance of the vessel to predict the speed of the sailing with the three statistical models: auto-regression, least squares estimates, and the maximum likelihood method. The authors find that wind speed and wave height have a considerable effect on vessel speed, both positive and negative. Meaning that the speed of the vessel can not only be reduced due to an increase in weather margins, it can also be increased. This is due to the fact that the direction of the wind and wave that correlates with the direction of the voyage can help “push” the vessel in the desired direction.

In order to increase the number of explanatory variables in the data, we collect data on weather variables from a third-party service. We collect data from the Copernicus Climate Change Service database of the European Union Space Programme, and match the coordinates of the vessel with the meteorological data at the corresponding coordinates.

3.3 Machine Learning

Machine learning is defined as the use of formal structures (i.e. machines) to make inferences (i.e. learning) (Clarke et al., 2009). Machine learning problems can typically be divided into two categories: supervised and unsupervised problems. In unsupervised learning, training data consist only of input values, and the objective of the algorithm is to extract meaningful patterns or structures from the data (Bishop, 2006). Supervised learning involves using a training dataset that consists of input examples paired with their corresponding target outputs. The algorithm aims to learn a mapping function that can accurately predict the target output for new observations (Gkerekos et al., 2019).

There is a growing research trend focused on artificial intelligence and machine learning techniques to improve fuel efficiency in ships, predict sailing speeds, and route selection (Zis et al., 2020). Many studies in this field rely on the use of artificial neural networks (ANN) and other machine learning methods to forecast fuel consumption in various sea conditions, with the aim of optimizing voyages to reduce fuel usage or ensure timely arrivals (Wang et al., 2016; Zheng et al., 2019; Du et al., 2019; Mao et al., 2016; Beşikçi et al., 2020). These models have been shown to be effective in predicting fuel, cost, and emission savings. And, as mentioned above, recent studies have even used ANNs to model

and predict port congestion (Abualhaol et al., 2018a; Peng et al., 2022).

Deep Neural Networks (DNN) are frequently used to predict road traffic flow, with a variety of models being applied, including combinations of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) (Wu et al., 2020; Bogaerts et al., 2020). Due to the time series characteristics of the AIS data, Peng et al. (2022) applied an RNN as their prediction model. Furthermore, due to its proven efficiency in predicting road traffic congestion (Mohanty et al., 2020; Shin et al., 2020), an LSTM RNN approach was implemented.

LSTMs were originally created to address the challenge of long-term dependencies, which hinder the effectiveness of traditional RNNs due to issues such as exploding and vanishing gradient problems (Hochreiter and Schmidhuber, 1997). The finished model of the study was a stacked multilayer LSTM model with three layers. This model structure is the most widely used and is known to make reasonable predictions (Greff et al., 2017). Abualhaol et al. (2018b) also used an LSTM RNN model for their well-performing prediction model of port congestion in the port of Singapore. The researchers opted to use a stacked multilayer model consisting of five layers.

3.4 Performance Metrics

To measure the accuracy of the model's predictions, we need to select performance metrics. However, no performance metric is optimal in all cases, as each metric has its drawbacks and advantages. Therefore, it is important to choose the performance metric based on the purpose of the prediction and the nature of the model (Swalin, A., 2007). Therefore, a selection of several metrics will be useful. Our chosen metric selection consists of some of the most commonly used ones: the root mean squared error (RMSE), the mean absolute error (MAE), and the symmetric mean absolute percentage error (sMAPE).

RMSE is an absolute measure of the fit (Abebe et al., 2020). It shows the proximity between the values of the estimation model and the values of the observed data, and is the square root of the variance (i.e. standard deviation) of the individual differences in the residuals. Abebe et al. (2020) argue that for models that are developed with the purpose of predicting, this measure is an appropriate choice when evaluating the accuracy of the model's predictions. On the contrary, Abebe et al., Willmott and Matsuura (2005)

argues that RMSE is a poor measure of prediction error. The main drawback of RMSE according to Willmott and Matsuura, is that the impact of each error on the total error is proportional to its square. Consequently, larger errors have a relatively stronger impact on overall error compared to smaller errors. The mathematical definition of RMSE is shown below.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.1)$$

Willmott and Matsuura (2005) recommend the use of MAE instead of RMSE. MAE is a straightforward, yet efficient metric to quantify the expected errors of the predictions. It does this by adding the absolute differences between the predicted and actual observed values before dividing by the number of samples. Therefore, a lower MAE value is preferred, since it indicates a lower expected prediction error. The mathematical definition of MAE is shown below.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.2)$$

When reviewing comparable research with the same objective as this thesis, we see that RMSE and MAE are preferred measurements of predictive precision (Peng et al., 2022; Abualhaol et al., 2018b). A drawback of MAE is that it is difficult to interpret when comparing models on data of different scales. MAPE is trying to overcome this by measuring it as a percentage. Although MAPE is scale-independent and easier to interpret, it still has its own drawbacks. According to Gkerekos et al. (2019), the error metric favors underestimated predictions, since there is a maximum limit of 100% error for such predictions, while there is no corresponding limit for overestimated predictions. These issues are solved by sMAPE which sets the upper limit to 200% and treats the actual observation values of zero as equal to the upper limit. These adjustments make the sMAPE a comparable but improved variant of the MAPE. The mathematical definition of sMAPE is shown below.

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \quad (3.3)$$

4 Data

This section describes the methodology for data acquisition, preprocessing, and feature selection. Our data processing approach is similar to the workflow demonstrated by Abebe et al. (2020). A detailed illustration of the data methodology can be found in Appendix A1.

4.1 AIS Data

The AIS provides high-frequency real-time positioning data on almost all commercial vessels. Originally, AIS was developed with the purpose of avoiding ship collisions and for governmental monitoring (UN Statistics, 2020), but over time it has become a powerful tool that not only provides operational decision makers with valuable real-time information, but also gives researchers the possibility of studying maritime trade processes and flows in detail. Moreover, collecting AIS data was initially a challenging process due to being highly regionalized. Communication between ships or between ships and the coast using AIS was initially limited to the VHF radio wave range, which provided coverage of only 10-20 nautical miles, thus imposing constraints on direct communication distances. However, from 2008, AIS data transmission has improved worldwide, as satellites with AIS receivers are now able to receive data transmitted by ships (Yang et al., 2019).

There are two types of AIS transceivers, Class A and Class B, which differ in their reported data fields and reporting frequencies. A Class A transceiver from a ship transmits information in 15 data fields, classified as static, dynamic, and voyage-related (Table 4.1). Dynamic information is automatically sent by Class A transceivers every two to ten seconds when the ship is moving and every three minutes when it is anchored. Class A transceiver transmits static and voyage-related information every six minutes, regardless of the ship's navigational status. In contrast, Class B transponders transmit a reduced set of data, excluding the IMO (International Maritime Organization) number, draft, destination, ETA, rate of turn, and navigation status, and the frequency of these reports is a minimum of five seconds.

Although there are many positives with AIS data, there still exist some limitations and drawbacks with the data itself. The large number of errors and inaccuracies that can be

Table 4.1: AIS Data Fields

Static	Dynamic	Voyage Related
MMSI	Vessel Position	Vessel Draft
Call Sign and Name	Position Time Stamp in UTC	Hazardous Cargo (Type)
IMO Number	Course Over Ground (COG)	Destination
Length and Beam	Speed Over Ground (SOG)	
Type of Vessel	Navigational Status	
Location of Antenna	Rate of turn (ROT)	

found result from several of the variables being manually reported (Yang et al., 2019). Data that are manually reported in AIS are static information, such as MMSI (Maritime Mobile Service Identity), vessel length and width, IMO number, name, type, call sign, and voyage-related information, such as estimated time of arrival (ETA), draft, and intended destination. As a result, missing or inaccurate ETAs have frequently been reported (Baldauf et al., 2008; Watson et al., 2015). Moreover, Baldauf et al. (2008) also found other inconsistencies in other AIS variables. Therefore, it is important to process the data accordingly and to consider the implications of erroneous information when working with AIS data.

4.2 Meteorological Data

Based on its effect on speed, route and risk, meteorological and oceanographic data are included in our prediction model. Historical data are retrieved from the Copernicus Climate Change Service database of the European Union Space Programme, which provides easy-to-access data on past, present, and future forecasts of weather and ocean parameters. More specifically, we use the ERA5¹ hourly data on single-level datasets, which contain historical data from 1979 to the present.

ERA5 data at the single level consist of parameters such as wind speed and direction, wave height, current and swell direction, temperature, and precipitation. Studies have shown that wave and wind conditions exert the most significant influence on ships (Carlton, 2018), while current and swell conditions have less influence (Adland et al., 2020; Abebe et al., 2020). Therefore, we only include wind speed and direction, wave height, and local precipitation as input variables in our prediction model.

¹European ReAnalysis, 5th generation.

The reason for including local precipitation is to capture the effect of rainfall in the region on soybean production, as rainfall is the main factor responsible for fluctuations in agricultural yield (Carmello and Sant’Anna Neto, 2016). These precipitation data are collected in Lucas do Rio Verde, the center of Brazil’s highest-producing soybean region (Martinelli et al., 2017). Precipitation data is also collected in the port of Paranaguá because rain has a disruptive effect on port operations because heavy rain can cause visibility problems for both ship crews and stevedores, as well as flood problems. As late as 2022, the port of Paranaguá was cut off from road and rail access due to flooding caused by heavy rain (Mano, 2022).

More details about the variables and the necessary feature engineering will be covered in Section 4.4.2.

4.3 Additional Data

In addition to the AIS and meteorological data, we want our prediction model to include other types of data that could influence the traffic of vessels in and around the port of Paranaguá. This is information on vessel-specific characteristics, and financial and agricultural data such as soybean, bunker oil, and fertilizer price.

Information on the vessels themselves, such as deadweight tonnage, size, and draft, will be included in the model and is retrieved from the Clarksons’ World Fleet Register database. Vessel statistics are collected for individual vessels in the final AIS dataset. The reasoning behind including this information in the model is that the size of the vessel will affect the operational capabilities of the port, as larger vessels need more space and might occupy a larger number of available cranes for a longer period of time. The deadweight tonnage and draft of a vessel will influence the possibility of accessing the berth area, as well as the time spent at the berth. This is due to the variability in the water level in the port area along with these vessel specifications, which will decide the accessibility of the port for the individual vessel.

The cost of shipping is highly dependent on the price of the fuel that the ship uses, known as bunker oil, bunker fuel, or marine fuel (Notteboom and Vernimmen, 2009). We make the assumption that an increase in bunker oil prices, resulting in higher costs for the company, can impact the decision to sail immediately or delay, as well as the speed of

the voyage, subsequently affecting port traffic. This is why we choose to include the price of bunker oil as a predictor in our model. Historical daily data on bunker price was retrieved from Bloomberg. As the price of bunker oil can vary between locations, we chose to retrieve data from the price in the Rotterdam port, as this port is the second largest bunkering port in the world (Wijaya, 2023).

Brazil is one of the world's largest soybean producers (Statista, 2023), supplying more than 50% of soybeans in the world according to the US Department of Agriculture (2023). The price of soybeans reflects the potential earnings of the producers in Brazil, and we assume that they will be inclined to produce and sell more when their earnings are higher, leading to increased activity in the port. Historical data on soybean prices in Brazil are gathered from NASDAQ, which provides historical and real-time prices for financial assets and commodities (NASDAQ, 2023).

Lastly, we include weekly data on fertilizer prices in Brazil. This is done in order to capture how changes in production volume, either random or seasonally, might affect port activity. We believe that such changes will be captured by the fertilizer price because agricultural produce is the largest export in the Paranaguá port, and fertilizer plays a crucial role in agricultural production. Weekly historical data on the local price of fertilizer in Brazil are retrieved from Bloomberg.

4.4 Data Preprocessing

The scope and objective of this thesis are clear: to build a port congestion prediction model in the Paranaguá port in Brazil using AIS data with vessel-specific information, selected meteorological and oceanographic data points, as well as agricultural and financial data. When including many data sources, it is essential to preprocess the variables so that they represent reality and can be used in the finished model.

4.4.1 Processing AIS Data

AIS data were retrieved for the period January 2013 to June 2019, which was the period of time to which we had access from all data sources. Initially, the data included all vessels in the world for that period. Therefore, we filtered out all observations except those who either had Paranaguá as their destination or reported that their current location was in

the port area. Furthermore, we filter to keep only vessels classified as bulk carriers, as these are the only types of vessel in the data set that visit the main port of Paranaguá. There is also a large amount of container traffic in the port (Ward, 2016) but we lack data on these types of vessel and there is little overlap in shared resources between the two types of vessel, so we believe that omitting container vessels in our analysis will have little impact.

There may be some spillover effect of traffic from nearby ports in periods of high port traffic. However, the port of Itajaí, located just south of Paranaguá, mainly handles container cargo such as frozen foods, machinery and other equipment, so any spillover from Itajaí is less likely. The port of São Francisco do Sul, also located south of Paranaguá, is a large trading port for cellulose pulp and fertilizer, much like Paranaguá, creating an opportunity for spillover to occur. However, this thesis will not include an analysis of potential spillover and its influence on port congestion in Paranaguá.

The estimated time of arrival (ETA) is a variable in AIS that is known to often be inaccurate, as previously described (Baldauf et al., 2008; Watson et al., 2015). We also found this to be the case for our data, with missing values, ETAs reported to be many days in the past, and other unreliable reported arrival times (Figure 4.1). Therefore, we opted to disregard the reported ETA values and instead calculate the estimated arrival time manually. We did this using the Dijkstra algorithm, which is one of the most widely used pathfinding algorithms, and a detailed description of this calculation is found in the methodology section.

Furthermore, when examining the AIS data further, we noticed inaccuracies with some IMO numbers. The IMO number is a unique seven-digit number that is assigned to propelled, seagoing merchant ships of 100 gross tonnage and above with some exemptions (pleasure yachts, war ships) (International Maritime Organization, 2023). Observations with IMO numbers that were either of the wrong length, had the same number through all digits, or had missing values were discarded.

Lastly, we need to acknowledge the time irregularity in the AIS data. As the data are collected from different sensors and are broadcast from the vessels at a variable rate, and a differing number of ships being in range of AIS stations, the resulting observations are not evenly spread out on the time axis. Unevenly spaced time series has fewer

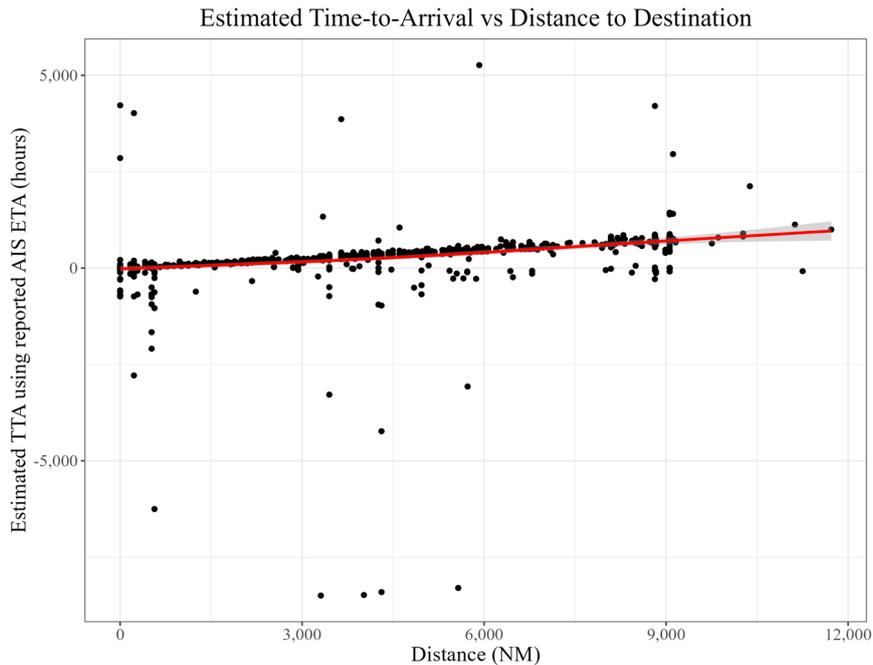


Figure 4.1: Distance (nautical miles) and estimated TTA using reported AIS ETA. TTA is calculated as the difference between the start of a trip and reported ETA. Note that all negative values are faulty.

methods for analysis and is harder to predict than regularly spaced time series. Therefore, time regularity is an important precondition for most common time series prediction models (Weerakody et al., 2021). Usually, the approach used to predict uneven time series is to transform the data into a regularly spaced time series, either by interpolation or downsampling, which causes bias and loss of information (Eckner, 2012). Manually downsampling the data to the extent of having the same number of observations per day would be detrimental to the model’s practical objective and use case; therefore, we keep the data and observations as they were originally. Newly developed methods, such as LSTM, combined with increased available computing power, have made it feasible to predict large, unevenly spaced time series in their unaltered forms.

Since there is no constant interval between observations, predicting a specific time horizon is impossible. Rather, one would predict a specific number of observations. Since the objective of this paper is to make a one- to three-month forecast, we need to find a proxy for that amount of time. The average number of observations per month is 1,233 or roughly 1,200, so the number of observations to predict to forecast three months would be $1,200 \times 3 \text{ months} = 3,600$ observations.

4.4.2 Weather Data Encoding

Meteorological and oceanographic data was retrieved and matched with the corresponding points in space and time for the vessels in the cleaned AIS dataset. Wind data was transformed from using the northward and eastward component system into separate variables for direction and strength. Wind data follows the meteorological convention where the data are divided into two components, u which is positive for a west-to-east flow (eastward wind) and v which is positive for a south-to-north flow (northward wind) (Blanchonnet, 2018). We hypothesize that the presence of wind is a bigger factor than the specific direction of the wind in whether ships will need to enter port. Therefore, by transforming the two directional components into a speed and direction component, we make the variables easier to interpret for humans. The transformation is done using the following equations:

$$V_s = \sqrt{u^2 + v^2} \quad (4.1)$$

$$V_d = \text{atan}_2(-u, -v) \times \frac{180}{\pi} + 180 \quad (4.2)$$

Equation 4.1 returns the wind strength in meters per second. Equation 4.2 returns the direction of the wind in the range of 0-359 degrees.

The observations of the wind and wave height data are matched to the observations of the ships by finding the closest match by date and geographical haversine distance. Observations of precipitation are matched with the rest of the observations in the dataset by date.

5 Methodology

5.1 Spatiotemporal Data Mining

5.1.1 Vessel Turnaround Time

LSTM is a supervised model, which means that it needs a labeled dataset where the output value is already known. Since we do not directly possess a variable for the level of port congestion, one must be synthesized from the other variables. To be able to quantify the level of port congestion, we first need to define port congestion. Waiting time is a common indicator used by online services that monitor current port congestion, used by industry leaders such as Marine Traffic (Marine Traffic, 2023) and also in academic research articles (Yeo et al., 2007). The waiting time is also easy to interpret for the ship operators.

Rather than using the vessel's waiting time at anchor before they are allowed to enter the berth area, we want to use a measure of congestion that will capture the entire port visit. In addition, ship operators most often go directly to berth when arriving at the port. Therefore, we adapt a similar congestion measure to Peng et al. (2022) and Bai et al. (2022), using the vessel turnaround time as our port congestion measure.

We create an algorithm that calculates how long it takes a vessel to visit the port. Two geofence areas are created by manually creating polygons in the areas where the vessels are observed anchoring and berthing (Figure 5.1). A more dynamic approach would use a clustering algorithm to dynamically identify these areas. A complete visit is defined as entering the anchorage zone from the open sea, entering the berthing area, and then leaving the port. The turnaround is averaged per day from the day it entered the anchorage area, i.e. if only one ship entered the anchorage area on the first of January and left the port on the third, the average turnaround time for the first of January would be two days.

The algorithm generates a time series with some gaps on days when no vessels were observed to make a full visit. These values are filled in with linear interpolation. A moving average with $n = 15$ days is created to reduce the jaggedness of the time series. The result can be seen in Figure 5.2 below, where the top panel shows the moving average time series.

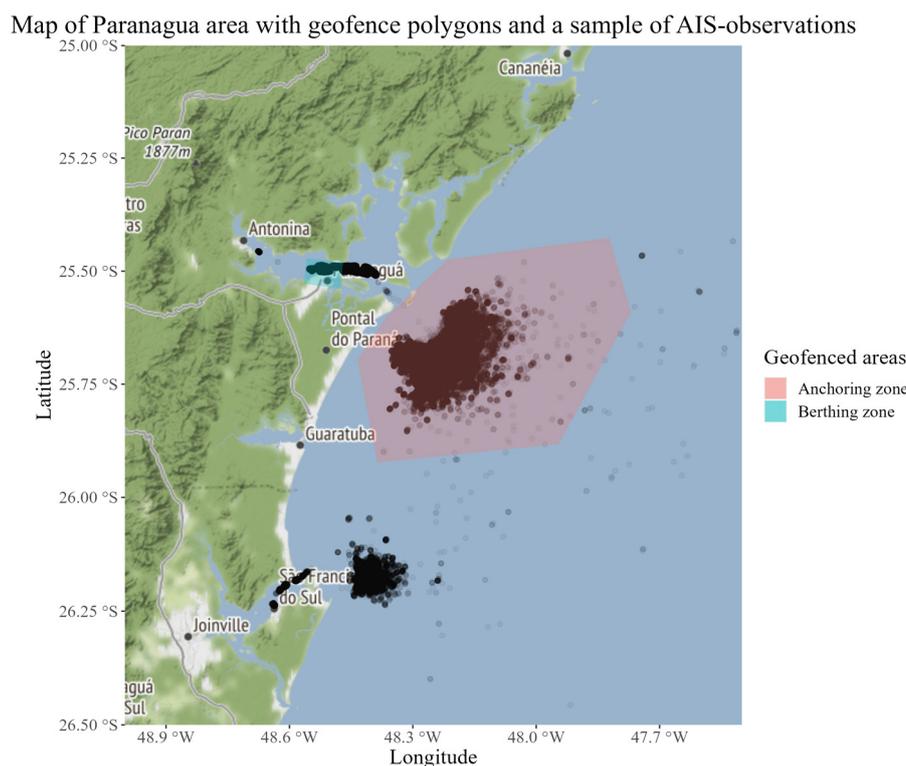


Figure 5.1: A map of the Paranaguá port area with the geofenced areas of interest; the berthing zone where vessels are in port, the anchoring zone where vessels are waiting to enter the port, and the navigation zone where vessels are moving between them. Also on the map is the port of São Francisco do Sul with corresponding anchoring cluster, south of Paranaguá.

The other four panels below show the same time series but are split using seasonal and trend decomposition using Loess² (Cleveland et al., 1990). It is decomposed additively, which means that adding the four components together results in the full time series. The first component is the trend, which shows that there was a considerable reduction during 2016 but that it was relatively stable afterwards. The second component, the seasonality, shows that there seemingly is a pattern that repeats over the year-cycle with peaks around New Year's and a lower turnaround time during the summer months. This corresponds to some extent with the seasonality of soy production in Brazil, with December and January experiencing the peak of the rainy season and soybean exportation beginning in early February (Cordonnier, 2021). The third component is the intra-week seasonality, showing a pattern of fluctuations with a turnaround time on Sundays and Mondays lower than the rest of the week. The last component is the remainder, which is all the fluctuations that the previous components cannot explain. The magnitude of the remainder is roughly

²Locally Estimated Scatterplot Smoothing

the same as the yearly seasonality, which means that there are significant unexplained variances.

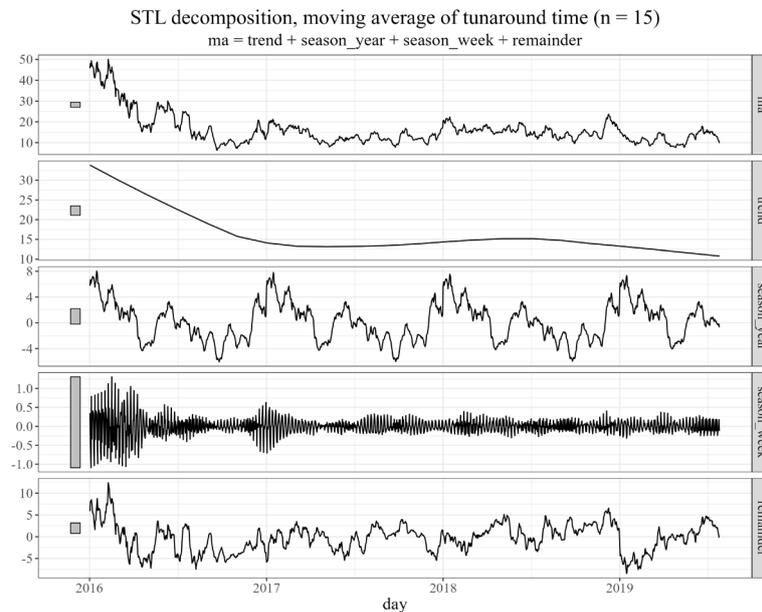


Figure 5.2: An STL-decomposition of the last three years showing the linear components of turnaround. The full turnaround can be seen in the top panel.

Finally, we now have the label value that our model will try to predict. This dependent variable is the turnaround time and is calculated as the time from the first time the vessel enters the geofence area of the port until it leaves the same area.

5.1.2 Synthesizing ETA with Dijkstra's Algorithm

Knowing when ships will arrive at the port is essential information to predict congestion. The AIS dataset that we have access to has an Estimated Time to Arrival (ETA) variable. However, as explained in Section 4.4.1 the reliability is sub-par due to the manual input nature of the variable. Therefore, we need to calculate our own ETA for each observation to replace the unreliable ETA variable. To accomplish this, we must use a pathfinding algorithm, as the time required to reach Paranaguá varies significantly depending on where the vessel is and how much land is needed to be navigated past. Eurostat has already created a Java library called SeaRoute for the purpose of calculating the shortest route over water between two coordinate points, which we utilize. This library is based on Dijkstra's algorithm, an algorithm for finding the shortest path between two nodes in a weighted graph (Dijkstra, 1959). The distance is used to calculate a new value for ETA

by multiplying it by the average steaming speed in the dataset.

In our dataset, no bulk carriers were observed passing through the Panama canal when traveling to Paranaguá. For one of the more common routes in our dataset where going through the canal would be faster; anywhere from the North American west coast to Paranaguá, the canal route would only be 600 nautical miles shorter, translating to roughly two days of sailing. The canal, on the other hand, will often take longer due to waiting times in addition to incurring significant passage costs. Therefore, we opted to remove the ability for vessels to transit through the Panama canal when calculating the ETA.

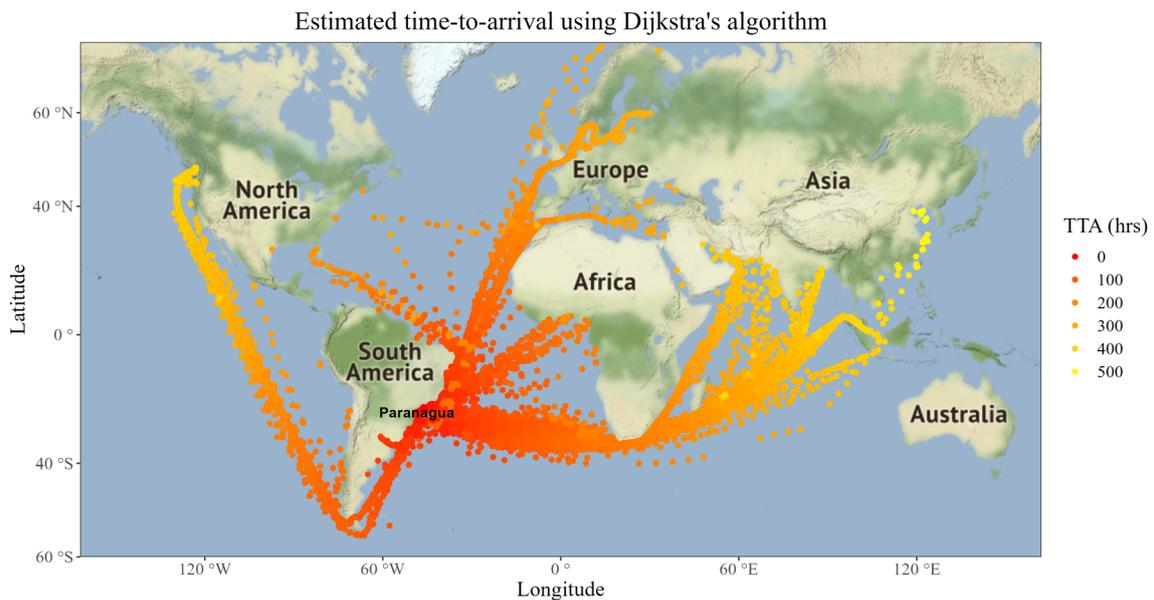


Figure 5.3: A map of ships traveling to Paranaguá color-coded to show estimated time-to-arrival.

The ETA variable itself is not useful for the LSTM network, as it needs the current timestamp of the observation as well to make it relevant to measure when the vessel will arrive. Therefore, the ETA variable is transformed into a time-to-arrival (TTA) variable by taking the difference between the ETA and the current timestamp of the observation.

5.2 LSTM Neural Network

The LSTM neural network is an advance of traditional RNN (Hochreiter and Schmidhuber, 1997). The main difference between LSTM and RNN networks is that LSTMs have a more complex architecture with a memory cell and gates that allow them to selectively

remember or forget information. On the other hand, RNNs have a simple structure where each output is dependent on the previous output and a hidden state, which is passed from one time step to the next. However, the problem with traditional RNNs is that they suffer from vanishing or exploding gradients, making it challenging to capture long-term dependencies. The vanishing or exploding gradient problems are a common issue that can occur during the training of deep neural networks. It happens when gradients become very small or very large as they propagate backwards through the network during training, which can result in slow training or even failure to converge. On the other hand, LSTMs are designed to address the problem of vanishing gradients by incorporating a memory cell, which allows them to remember or forget information at each time step selectively. This makes LSTMs better suited for tasks that require the network to capture long-term dependencies and remember past information. Therefore, it is more suitable for time-series prediction, particularly for time-series data that exhibit seasonality.

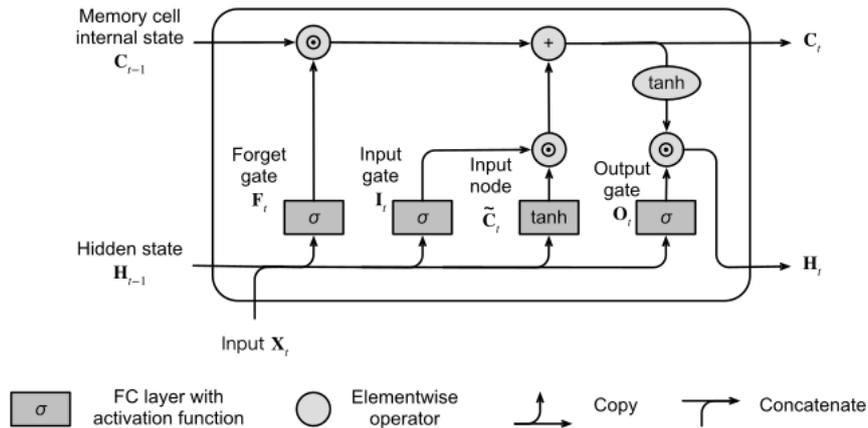


Figure 5.4: A diagram showing the structure of a long short-term memory neural network.

The basic components of an LSTM network are three gates and a cell state (Olah, 2015). The forget gate f_t , takes the previous hidden state, h_{t-1} , and the current input x_t as input, and using a sigmoid layer, determines the amount of the previous cell state, C_{t-1} , to "forget" by returning a value between zero and one.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5.1)$$

Next, the model will decide what information will be stored in the cell state. This is determined by the input gate, i_t , and the input node, \tilde{C}_t . The input gate using a sigmoid

layer determines the values that are to be updated. The input node creates a vector of new candidate values that can potentially be added to the cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5.2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (5.3)$$

The cell state from the previous cell is updated with the values from the forget gate and the input gate. Subsequently, the cell state is multiplied by the forget value f_t and then updated by adding $i_t * \tilde{C}_t$, which is the vector of candidate values scaled by how much the input node decided to update each state value.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5.4)$$

Next, the hidden state, or more simply the output, h_t , is set by the output gate o_t , which is a sigmoid layer that determines which parts of the cell state to output. The cell state goes through a tanh function, which reshapes the cell state to have a value between -1 and 1. Finally, the weight of the sigmoid layer is multiplied by the output of the tanh function, and the result is the new hidden state.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5.5)$$

$$h_t = o_t * \tanh(C_t) \quad (5.6)$$

5.3 Multi-Step Multivariate LSTM Model

The objective of this thesis requires us to build a multi-step multivariate model. This is a model that predicts more than one unit of time ahead (multi-step), and takes in more than one independent variable as input (multivariate) (Pang, 2020).

Creating a single prediction model to predict any desired horizon would be ideal. However, such a model is unfortunately so timely and computationally demanding that it is simply not possible with the hardware and software to which we have access. Therefore, six models are built, each model being optimized to its own prediction span ranging from one

to twelve weeks. Each model uses an input matrix of two times the number of desired output values with 21 independent variables.

The models are built and trained using the Keras library in Python, which is an open source interface for the popular TensorFlow library built by Google Brain (Helms et al., 2018). The models are trained on a powerful workstation with 64GB of RAM and a CPU with 24 logical threads, each running at up to 4.6GHz. We decided to implement the model on a CPU instead of a GPU due to the relatively easier implementation complexity and ease of debugging. In addition, most consumer GPUs lack the necessary amount of RAM to implement larger models directly.

5.3.1 Data Preparation for LSTM

Before the data can be used as input in the LSTM model, it needs further preparation. First, the data must be divided into training, validation, and test sets. This is done to avoid overfitting (Aggarwal, 2015). We split the data so that the training set consists of all observations for the first 34 months of the dataset, the validation set for the following 6 months, and the test set consists of the remaining 17 months, corresponding to a 60%, 10% and 30% split.

Next in data preparation is feature scaling. Feature scaling is essential when working with data where the features have different ranges or units of measure. It is especially important to scale the data when working with machine learning models that utilize optimization techniques, as neural networks do. This aims to ensure that each feature contributes equally to the model and to avoid the dominant influence of features with higher values (Bhandari, 2020).

We scale our data by normalization, that is, scale the feature values to a range of 0-1. The mathematical formula for scaling using normalization is shown in Equation 5.7. Furthermore, we also transform the validation and test data using the minimum and maximum values of the features in the training set. In this way, we avoid potential data leakage during testing.

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (5.7)$$

We now have a final dataset that includes all the 21 desired input variables, as well as the turnaround time as the dependent variable. The final dataset consists of AIS data, Clarksons' World Fleet Register data, commodity price data, weather data, and self-generated variables. The complete data are shown in Table 5.1, including the number of observations and the time span of the data.

Table 5.1: Complete Dataset

AIS	Clarkson's	Commodity	Weather	Self-Generated
Length Overall	Gross Tonnage	Bunker Oil Price	Wind Speed	Turnaround Time
Beam	Deadweight Tonnage	Fertilizer Price	Wind Direction	Time To Arrival
Speed over Ground	Length	Soybean Price	Wave Height	Distance
Course over Ground	Age		Temperature at Ship Position	
	Draft		Precipitation Paranaguá	
			Precipitation Lucas do Rio Verde	
Number of observations:				67,578
Date span:				2013-04-29 - 2019-07-27

Data must be fed into LSTM models in a specific manner. First, the data are divided into input sequences two times the size of the desired output. Hence, when the desired prediction horizon is one week, corresponding to 300 data points, the length of the input sequence is 600 points. Then, multiple sequences are pooled, creating one batch. The optimal number of sequences that make up one batch, known as batch size, is determined by hyperparameter tuning, which is covered in Section 5.3.3. Lastly, LSTM models require the input data to be three-dimensional where the first dimension is the number of samples, the second dimension is the sequence length, and the third is the number of features.

LSTMs follow the sliding-window prediction principle. This means that one input sequence is used to predict one output sequence before sliding one data point and repeating the same procedure, creating overlapping sequences (Figure 5.5). In our case, for a one-week prediction horizon, the previous 600 data points were used to predict the next 300 points.

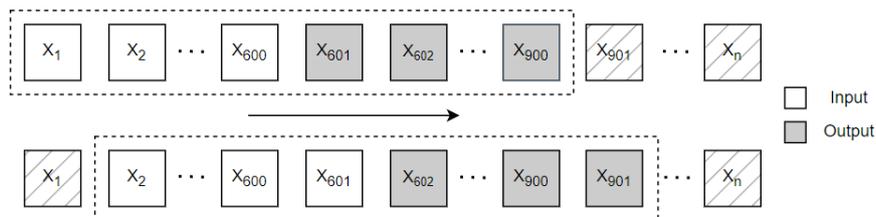


Figure 5.5: Sliding window technique for one week prediction. An input sequence of 600 data points is used to predict the output sequence of 300 points, before sliding one data point and repeating the process.

5.3.2 Model Training

During training, the neural network learns by updating its parameters using an optimization algorithm that adjusts the weights and biases of the nodes in the network. This process continues until the model reaches a point where a given loss measure stops improving. The loss measure is set by us and can be any performance measure used for regression problems. We set the loss measure for the training of our models to be the mean squared error (MSE), as it is a widely used measurement of the performance of the prediction model (Brownlee, 2019). This metric is only used to evaluate performance during model training.

Adam³ is an optimization algorithm commonly used in neural network training that uses adaptive learning rates and momentum to accelerate convergence (Kingma and Ba, 2014). It maintains past gradients and squared gradients to adjust the learning rate for each parameter and uses momentum to smooth out the optimization process. Adam is particularly robust to noisy or sparse gradients and has been shown to converge faster than other optimization algorithms on many deep learning tasks.

An epoch is a complete iteration through a dataset during the training phase of a model. During an epoch, the model is trained on each data point in the dataset and the weights of the model are adjusted to optimize the performance of the model (Chollet, 2021). Choosing the right number of epochs is important, as the model needs enough time to learn the patterns in the data to avoid underfitting. On the other hand, it must not be allowed to train for so long that it starts memorizing the training data, i.e. overfitting. Early stopping is a method where one uses a dynamic number of epochs, stopping only when the validation loss stops improving after multiple epochs. In our case, if training does not result in a lower loss for three consecutive epochs, the model stops training and uses the weight and biases from the best epoch.

5.3.3 Hyperparameter Optimization

The model training process identifies the fitting values for most parameters. However, some parameters cannot be directly learned during this process. These are called hyperparameters. Hyperparameter optimization, or tuning, is the problem of finding

³The name is derived from Adaptive Moment Estimation.

the fitting values for these parameters (Feurer and Hutter, 2019). The hyperparameters that need to be tuned to find their optimal values are the number of layers in the neural network, the number of units in each layer, the dropout ratio after each layer, and the batch size of sequences fed into the model.

In terms of the tuning itself, it is performed using the KerasTuner tuning library (O'Malley et al., 2019). There are several methods to perform parameter search, where the most used methods are Random search, Grid search, and Bayesian optimization. All have their own advantages, but we have opted for using the Random Search method because of its ability to discover hyperparameter combinations that are harder to see intuitively. Random Search uses a random selection process within a bounded domain of hyperparameter values and finds the optimal parameter values within the set domain (Brownlee, 2020).

5.4 Model Structure

After the hyperparameter tuning has been completed, the optimal network structure for each model is the result. These structures are shown in Table 5.2.

All models have the structure of three LSTM layers, where the first two layers are fed into a dropout layer with a dropout rate of 20%. A dropout layer is used in neural networks to prevent overfitting by randomly setting a fraction of the input units to zero during each training iteration, forcing the network to learn a more robust and generalizable set of features. Finally, this is fed into a dense layer that outputs a vector containing the number of points corresponding to the desired prediction span.

The larger the forecasting horizon, the more units the model should have in the hidden layers. Due to the limited computing power available to us, the larger models are too large to sufficiently learn the information in the data. The two largest models with a horizon of eighth and twelve weeks have fewer units in the hidden layers compared to the models with shorter horizons. However, they took more than four times the time to train and were left to run for multiple days on a professional workstation. Thus, one should not look at the performance of these models as proof of their inability to predict accurately for longer horizons, rather that the computational complexity increases severely with a longer forecasting horizon. Consequently, it was impossible to train the 12-week model, and we had no choice but to leave this model out and continue with only the 1- to 8-week

models.

Table 5.2: Structure of LSTM Models

Components	Models				
	1 Week	2 Weeks	3 Weeks	4 Weeks	8 Weeks
Steps In	600	1200	1800	2400	4800
Steps Out	300	600	900	1200	2400
LSTM1 Units	128	128	128	128	64
Dropout	20%	20%	20%	20%	20%
LSTM2 Units	128	128	128	128	64
Dropout	20%	20%	20%	20%	20%
LSTM3 Units	128	128	64	64	48
Dense Units	300	600	900	1200	2400
Batch Size	32	32	32	32	32
Epochs	2	4	5	5	5
Parameters	323,884	352,984	316,292	335,792	194,336
Run Time	2h:52m	8h:03m	12h:17m	12h:39m	25h:54m

The models were trained over different numbers of epochs in which the training ended after three consecutive epochs in which no improvement in the validation loss was observed. Then the weights of the best-performing model were restored and used in the trained model.

6 Analysis and Discussion

In this section, we present the models' predictions and evaluate their performance with the metrics discussed in Section 3.4. Additionally, we compare the models with two more simple prediction models, and we investigate feature importance. Finally, we discuss the implications of our results for ship operators and owners.

6.1 Model Predictions

When predicting values for the test set using the five models, the models produced vectors of predicted points with lengths of 300, 600, 900, 1,200 and 2,400, respectively. These lengths correspond to forecast horizons of approximately one week (300 points) to eight weeks (2,400 points) into the future. The predictions were made for the entire test dataset, which spans a period of roughly 16 months.

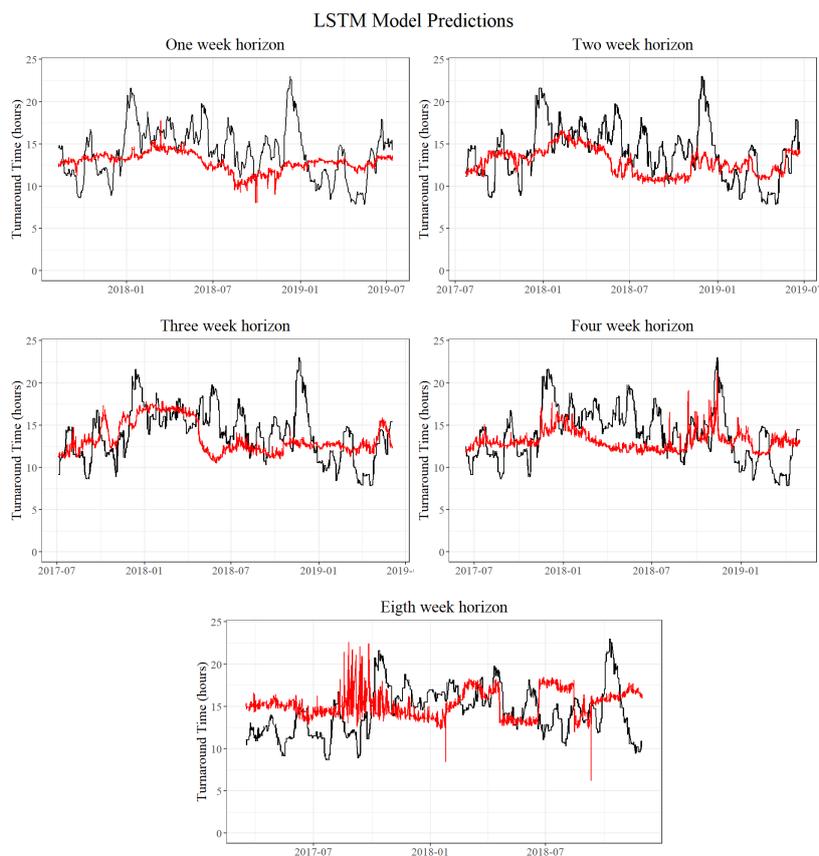


Figure 6.1: Predictions using LSTM with different forecast horizons on the test data. The length of the test set varies slightly due to the differing lengths of the input for each prediction point.

Figure 6.1 displays a comparison between the predicted and actual turnaround time values. Turnaround time is our measure of congestion, meaning that changes in the turnaround time, positive or negative, are proxies of changes in port congestion. For ship and port operators, this figure represents the duration spanning a vessel’s entrance to the port until its departure, providing valuable insights into traffic volume and the level of congestion experienced at the port.

6.2 Performance Evaluation

Our goal is to evaluate the performance of each model individually, taking into account its different prediction horizons. This allows us to understand how the accuracy of predictions changes as the horizon lengthens and predictions are made further into the future. As shown in Table 6.1, the prediction accuracy remains relatively consistent across different horizons. However, as expected, the models with the longest horizons exhibit a decline in performance, except for the one- and two-week prediction horizons, which, surprisingly, perform worse than the three- and four-week predictions horizons. This might be due to the shorter input sequence length, making it harder for the model to learn seasonal patterns and generate accurate predictions.

Table 6.1: LSTM Models Performances

Metric	Models				
	1 Week	2 Weeks	3 Weeks	4 Weeks	8 Weeks
RMSE	3.178	3.105	3.053	3.063	3.461
MAE	2.554	2.401	2.374	2.529	2.898
sMAPE	0.186	0.176	0.173	0.184	0.198

According to Table 6.1, the one-week model has an average prediction error of 18.6%, as measured by the sMAPE. In absolute terms, according to the MAE, the model’s average prediction error is 2.554 hours. This means that on average, the model’s predictions for turnaround time deviate from the actual values by 2.554 hours. Moreover, the best performing model is the three-week model, which generated predictions with an average error of 2.374 hours. Investigating the table further, we observe an improvement in performance up to the three-week model, where the trend reverses and the average error

in the turnaround time increases to 2.898 hours for the eight-week model.

Figure 6.2 better visualizes the performance of each LSTM model. The graph illustrates, through sMAPE, that performance is improved from the one-week model to the three-week model. Thereafter, the performance notably worsens.

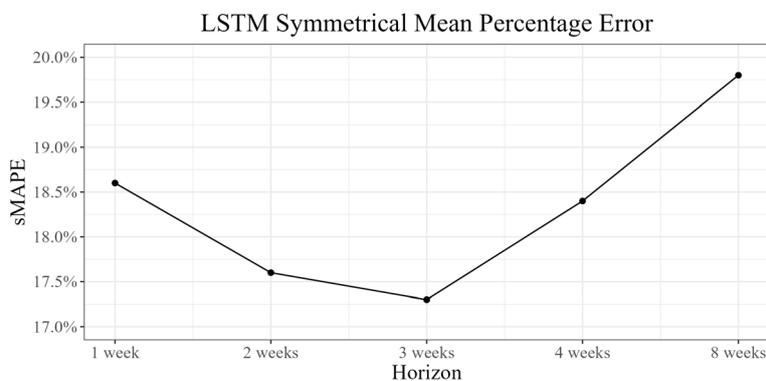


Figure 6.2: sMAPE for different forecasting horizons, one to eight weeks.

6.2.1 Comparing with Benchmark Models

To further assess the performance of our models, we compare their predictions with those of a benchmark model. While we use established metrics to measure the accuracy of the predictions of our models, it is also beneficial to compare their predictions with other simpler prediction models (Dhakal, 2017). A common benchmark model for evaluating more sophisticated models is the naïve model (Chyndman and Athanasopoulos, 2018). In addition to the naïve model, we also include a machine learning model that is more complex than the naïve model but less complex than the LSTM. The aim of this comparison is to establish whether our proposed method surpasses the performance of alternative models and to assess if the complexity of our model is justified for the problem at hand.

The benchmark models of our choice are a naïve average forecasting model and an XGBoost (Extreme Gradient Boosting) decision tree. The former model computes the average of all input sequences and uses the average as its prediction, while the latter fits an extreme gradient boosted decision tree where the final prediction is an average over all individual decision trees. XGBoost is a scalable algorithm based on the gradient boosting framework put forward by Friedman (2001), and has the reputation of being highly efficient in solving various types of data science problems (Qiu et al., 2021). It is an ensemble of decision trees where each new tree attempts to minimize the error of the previous one (Adland

et al., 2015). The benchmark predictions are shown in Table 6.2.

Table 6.2: Model Performance Comparison with sMAPE

Model	Horizon				
	1 Week	2 Weeks	3 Weeks	4 Weeks	8 Weeks
Naïve	0.583	0.583	0.583	0.583	0.583
XGBoost	0.685	0.682	0.924	0.819	0.682
LSTM	0.186	0.176	0.173	0.184	0.198

Table 6.2 displays the performance evaluation metrics for the naïve model and the XGBoost together with the LSTM. The comparison uses only sMAPE as it is a suitable and easy to interpret performance metric. Furthermore, sMAPE is the best metric to use when evaluating the degree to which predicted values align relative to actual observed values, regardless of the error in absolute terms (Kreinovich et al., 2014).

As shown in Table 6.2, our LSTM RNN outperforms the benchmark models. The performance of the benchmark models remains relatively consistent across different prediction horizons, similar to the LSTM RNN. Interestingly, as illustrated in Figure 6.3, the naïve average model performs better than the more complex XGBoost model. This suggests that it may be challenging for a simple machine learning model to accurately learn and capture the patterns in the data. As a result, using a simpler approach to predict the average observed value can lead to more accurate predictions.

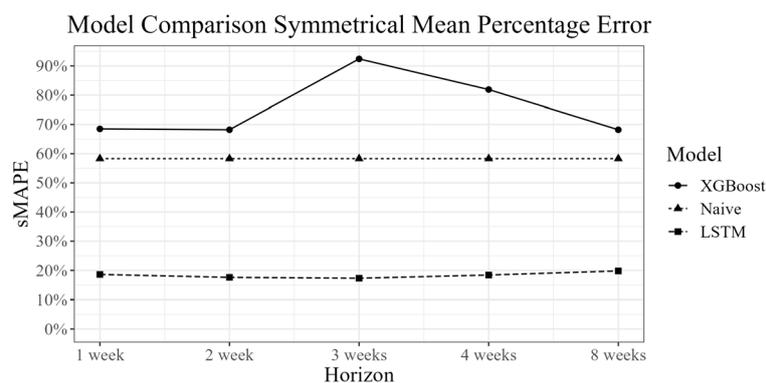


Figure 6.3: sMAPE for the XGBoost, Naïve model and LSTM with different forecasting horizons.

6.2.2 Feature Importance

Comprehending the rationale behind a model's prediction is of significant importance, comparable to the accuracy of the prediction itself. Analyzing feature importance in machine learning models is crucial for understanding the factors that drive predictions and for providing model transparency. SHAP⁴ is a powerful technique for conducting such an analysis (Lundberg and Lee, 2017) based on Shapley values⁵. By utilizing SHAP, the importance of each feature can be quantified, allowing for a comprehensive explanation of how each feature contributes to the model's predictions. SHAP provides insights into the relative influence of different features, highlighting the most significant contributors, and capturing potential interactions among them. This enables a deeper understanding of the model's decision-making process, aiding in model validation, interpretation, and trust-building.

We acquire the SHAP values using the SHAP library and its DeepExplainer (Lundberg and Lee, 2017), which is an algorithm to compute the SHAP values for deep learning models. Computing SHAP values is timely and computationally demanding. Therefore, we were unable to perform the computations on the original LSTM model but had to build a smaller yet similar model. This model has the same model structure with respect to layers, units, and input length as the smallest LSTM model, the one-week model. However, it differs in the number of output data points, only being 40 points compared to 300. As the model structure, input variables, and input length are identical, and the shap values seem unchanging over the time steps, we regard the acquired SHAP values as representative for the original model.

In Figure 6.4, we present a visual representation of the SHAP values of a subset of the test set. In this graph, positive SHAP values indicate an incremental increase in turnaround time, while negative values indicate the opposite effect. Also, the SHAP values represent the log-odds of the respective directional change in turnaround time occurring. Note that the SHAP values do not say anything about how different values within the variables are influencing the turnaround time, only that the variable itself has influence and to what extent. Although such exact influence is not obtained through the use of SHAP, it is,

⁴SHapley Additive exPlanations

⁵For which Lloyd Shapley received a Nobel Memorial Prize in Economic Sciences in 2012.

however, visualized through the feature values ranging from low to high.

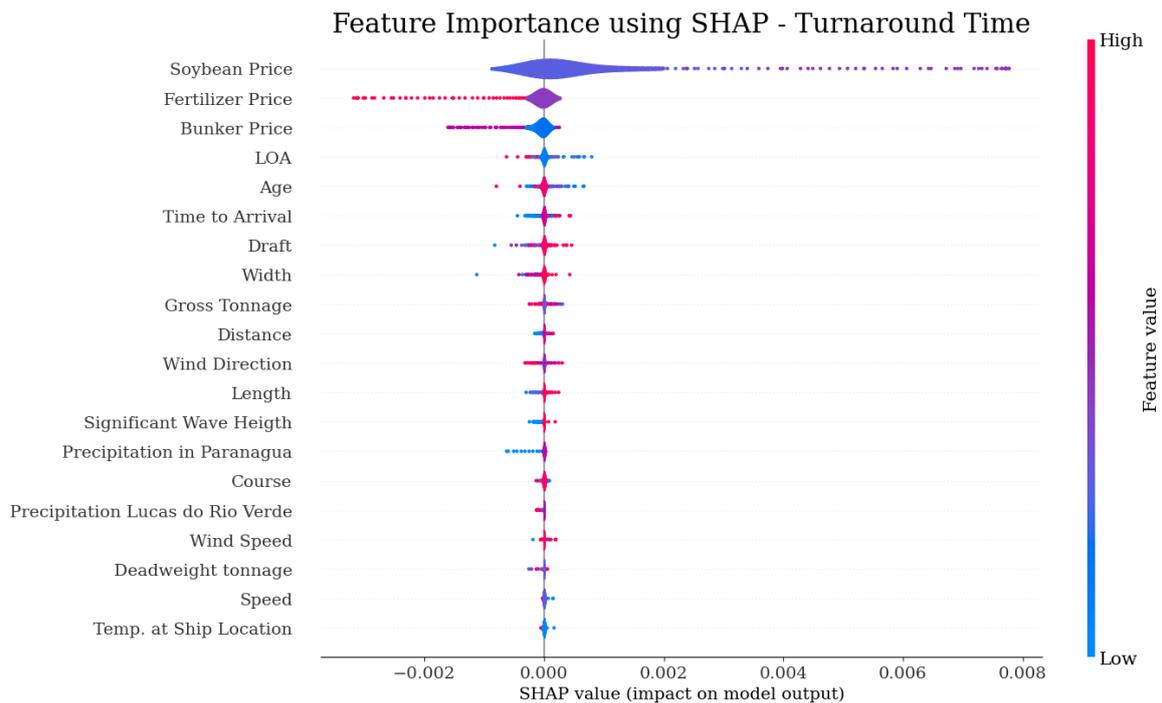


Figure 6.4: A beeswarm plot of SHAP values calculated for each variable using the SHAP library developed by Lundberg and Lee (2017).

The most influential feature that affects the turnaround time is the price of soybeans with a positive effect on the turnaround time, which means that the soybean price causes higher congestion levels. From the color in the variable, we see that an increase in the soybean price results in a longer turnaround time. Additionally, we note the importance of fertilizer price and bunker oil price as predictors. Both have an inverse correlation with the turnaround time, and the feature values suggest that higher prices results in lower congestion. This makes intuitive sense since higher prices mean higher costs for ship operators, as well as for producers of agricultural produce, making it less attractive to ship goods through the port and thus lowering volume. The significant importance of financial variables implies that there is a correlation between the commodity market and the congestion level in Paranaguá.

In addition, weather variables are not of great importance when predicting turnaround time. Wind direction and wave height show an unclear influence on the change in turnaround time, having both positive and negative effects. However, their influence is not great. Lastly, both precipitation in Paranaguá and the soybean producing region Lucas de Rio

Verde have a negative influence on turnaround time.

We acknowledge a weakness in the feature engineering of the wind direction variable. As it stands, the wind direction has values between 0 and 359 degrees, with values 0 and 359 having the same effect on a vessel from an operational point of view. The model does, however, view these values as complete opposites. A better way to format the wind direction variable would be to transform the values into sine and cosine values, which would range from -1 to 1 in a continuous manner. This method would make the variable more interpretable for the model.

6.3 Operational Implications

The findings of this study suggest that machine learning methods in the form of LSTM RNNs can bring valuable knowledge to ship operators and owners, when used in the context of port congestion. Being able to predict future port congestion in the port of destination gives more predictability for both ship and port operators. This, in turn, gives room for cost savings and risk adjustments throughout the vessel's journey.

The proposed models provide evidence that valuable insights can be extracted from the vast amounts of data generated in the shipping industry. With the appropriate analytical tools, patterns can be identified and used in operational decision making. There is, though, a high barrier to implementing a model like this in a production setting, as it requires large amounts of data, processing power, and man hours in the form of research, development, and maintenance.

More specifically, the operational implications of this study are that the proposed models are able to give shipping operators, such as G2 Ocean, estimates of what the future congestion will be at the destination port. The results show that for a prediction horizon of one week, the model predicts congestion with an average error of 2.554 hours (see MAE in Table 6.1). Furthermore, when expanding the prediction horizon to eight weeks, the model estimates future congestion with an average error of 2.898 hours. Visualization of the average error in turnaround time for different forecasting horizons can be seen in Figure 6.5.

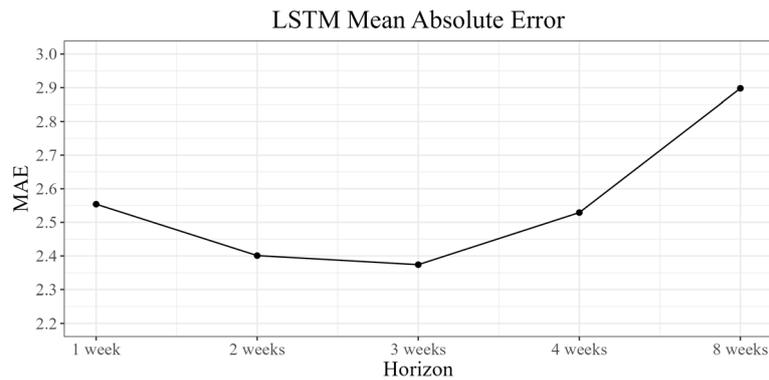


Figure 6.5: MAE for different forecasting horizons, one week to eight months. The MAE score is interpreted as the average error for predicted turnaround time in hours.

The reason we were not able to predict further into the future was due to the large amount of data and the large parameter space, and thus the model required too much computational power and time to find the optimal hyperparameters and weights. Lastly, the number of observations that are fed into the model might need to be higher for the model to find the long-term patterns required to predict into the future.

Despite the inability of the model to accurately predict port congestion beyond a few weeks in advance, it still has practical applications. While it may not be suitable for selecting contracts due to its limited predictive accuracy over longer time horizons, it can be used by ship operators to determine the optimal speed for their vessel, so as to arrive when a berth is available. Additionally, if permitted by their contract, the model can assist ship operators already underway in selecting which port to visit.

7 Limitations and Further Research

7.1 Limitations

As authors of this thesis, we acknowledge that there are limitations to our study. In particular, the lack of a publicly available database that contains congestion levels in the Paranaguá port forced us to construct our own measure of congestion as a dependent variable. This potentially weakens the validity of the results, as a computed measure of congestion might not measure the actual congestion with sufficient accuracy. This, of course, is impossible to be sure of, as no data is available on actual experienced congestion. Furthermore, using a self-defined congestion measure raises questions regarding the optimal methodology to quantify congestion and whether our measure is appropriate for this purpose.

To validate our findings, acquiring data on port activity directly from the port authorities in Paranaguá would be beneficial. Unfortunately, this is difficult in practice, as there is no consistent reporting system in place. A standardized reporting system that encompasses all ports would allow more comprehensive data collection, facilitate more robust analysis, and ultimately foster a more efficient and informed decision-making process for ship operators and port authorities alike. The existence of such a system would not only streamline operations, but would also provide valuable information for stakeholders across the maritime industry.

When considering the purpose of this thesis as constructing an accurate port congestion prediction model, it is obvious that the models built in this study are not dynamic enough nor include enough variables to be useful in other ports than Paranaguá. However, we would like to emphasize that the results produced by our models could serve as a proof of concept that it is possible to construct a universal real-time prediction model. Future researchers should therefore consider the performance of the LSTM RNN models put forward in this thesis and focus their attention on scaling up a similar model. The scope of this paper did not encompass the computational power required to train models with longer forecasting horizons.

7.2 Further Research

A generalization of the model that could dynamically identify berthing and anchorage zones would make the model much easier to adopt by ship operators, as well as minimize the possibility of human error. In addition, more data would be available to validate the performance of the model. Approaches for automatically identifying vessel clusters and ports using the density-based spatial clustering of applications with noise (DBSCAN) algorithm have been implemented by Bai et al. (2022), Fuentes (2021), Fuentes and Adland (2023) and Peng et al. (2022), but not in combination with a long forecast horizon.

In this thesis, the price of soybeans in Brazil was used as an input variable to capture the demand for soybeans. However, we recognize that this measure of demand is not optimal and that soybean crush margins would be a better choice. Unfortunately, we were unable to access any dataset on soybean crush margins. Therefore, future research should aim to include the soybean crush margin as a more precise measure of soybean demand, preferably the crush margin in China, since the country is the largest importer of soybeans from Brazil (Martinelli et al., 2017).

Furthermore, looking into the potential spillover effect from nearby ports on congestion, or at least including traffic volume at nearby ports in a prediction model, would be an interesting exercise for future studies.

Lastly, this study demonstrates the difficulties with long-term prediction models due to computational complexity. Future studies should therefore take note of the methodology used here and mitigate these barriers by utilizing a distributed computing cluster with dedicated GPUs or TPUs.

8 Conclusion

In this thesis, we have developed five long short-term memory recurrent neural network models which have been applied to predict future port congestion in Paranaguá port, over a prediction horizon of one to eight weeks. The computational complexity of the prediction problem forced us to limit the prediction horizon to eight weeks. Additionally, prediction models used information mined from AIS data, vessel characteristics, weather data, and commodity price data as input variables. Furthermore, we compared the prediction results from these LSTM models with two less complex models; a naïve average and an XGBoost model.

To our knowledge, this paper marks the first application of LSTM RNNs for long-term port congestion prediction. Furthermore, this study contributes to the existing literature by paying more attention to the importance of weather and other variables with causal effect on the development of port congestion. Our models demonstrated predictions with a satisfactory level of accuracy to be of practical use in shipping and port operations. In particular, the models outperformed two simpler models in terms of predictive performance. They demonstrated precise predictions for up to one month and showed the potential to predict further. However, they were limited by the computational complexity of the task.

The findings of this study have operational implications for ship operators in terms of optimizing ship speed and port calls. The proposed prediction models exhibit a growing average error the more data points are predicted, meaning that as the prediction horizon extends, the error of the forecast grows, highlighting the need for careful consideration in long-term predictions. The findings show that it is possible to predict congestion one week ahead with an average error of 2.55 hours, 2.40 hours for two weeks, 2.37 hours for three weeks, 2.53 hours for four weeks, and 2.89 hours for the eighth week. Therefore, we conclude that we succeeded in predicting port congestion weeks in advance with reasonable precision.

In conclusion, the predictions generated by the proposed models are shown to be promising with a satisfactory level of accuracy. The final evaluation of the presented models is that they serve their purpose and fulfill their objective within the constraints set by the authors and its inherent limitations.

References

- Abebe, M., Shin, Y., Noh, Y., Lee, S., and Lee, I. (2020). Machine learning approaches for ship speed prediction towards energy efficient shipping. *Applied Sciences*, 10(7):2325.
- Abualhaol, I., Falcon, R., Abielmona, R., and Petriu, E. (2018a). Data-driven vessel service time forecasting using long short-term memory recurrent neural networks. *2018 IEEE International Conference on Big Data (Big Data)*, pages 2580–2590.
- Abualhaol, I., Falcon, R., Abielmona, R., and Petriu, E. (2018b). Data-driven vessel service time forecasting using long short-term memory recurrent neural networks. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2580–2590. IEEE.
- Adland, R., Cariou, P., and Wolff, F.-C. (2020). Optimal ship speed and the cubic law revisited: Empirical evidence from an oil tanker fleet. *Transportation Research Part E: Logistics and Transportation Review*, 140:101972.
- Adland, R., Jia, H., Lode, T., and Skontorp, J. (2015). The value of meteorological data in marine risk assessment. *Reliability Engineering & System Safety*, 209.
- Aggarwal, C. C. (2015). *Data Mining*. Springer Cham.
- Ando, H. (2014). How we use fleet performance tools to increase our energy efficiency.
- Bai, X., Jia, H., and Xu, M. (2022). Identifying port congestion and evaluating its impact on maritime logistics. *Maritime Policy & Management*, pages 1–18.
- Baldauf, M., Benedict, K., and Motz, F. (2008). Aspects of technical reliability of navigation systems and human element in case of collision avoidance. In *Proceedings of the Navigation Conference & Exhibition, London, UK*, volume 28.
- Beşikçi, E. B., Arslan, O., Turan, O., and Ölçer, A. I. (2020). An artificial neural network based decision support system for energy efficient ship operations. *Transportation Research Part C: Emerging Technologies*, 66:393–401.
- Bhandari, A. (2020). Feature engineering: Scaling, normalization, and standardization.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag.
- Blanchonnet, H. (2018). What is the direction convention for the u and v components of winds?
- Bogaerts, T., Masegosa, A. D., Angarita-Zapata, J. S., Onieva, E., and Hellinckx, P. (2020). A graph cnn-lstm neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies*, 112:62–77.
- Brownlee, J. (2019). How to choose loss functions when training deep learning neural networks.
- Brownlee, J. (2020). *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery.
- Carlton, J. (2018). *Marine propellers and propulsion*. Butterworth-Heinemann.

- Carmello, V. and Sant'Anna Neto, J. L. (2016). Rainfall variability and soybean yield in paran state, southern brazil. *International Journal of Environmental & Agriculture Research*, 2(1):86–97.
- Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster.
- Chyndman, R. and Athanasopoulos, G. (2018). *Forecasting, Principles and Practice*. OTexts.
- Clarke, B., Fokoue, E., and Zhang, H. H. (2009). *Principles and theory for data mining and machine learning*. Springer Mathematics and Statistics eBooks.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition. *J. Off. Stat*, 6(1):3–73.
- Cordonnier, M. (2021). Annual month-by-month crop production cycle in brazil.
- Dhakai, C. P. (2017). A nave approach for comparing a forecast model. *International Journal of Thesis Projects and Dissertations*, 5(1):1–3.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271.
- Du, Y., Meng, Q., Wang, S., and Kuang, H. (2019). Two-phase optimal solutions for ship speed and trim optimization over a voyage using voyage report data. *Transportation Research Part C: Emerging Technologies*, 122:88–114.
- Eckner, A. (2012). A framework for the analysis of unevenly spaced time series data. *Preprint. Available at: http://www.eckner.com/papers/unevenly_spaced_time_series_analysis*, page 93.
- Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (1996)*, pages 226–231. Elsevier.
- Feurer, M. and Hutter, F. (2019). Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges*, pages 3–33.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.
- Fuentes, G. (2021). Generating bunkering statistics from ais data: A machine learning approach. *Transportation Research Part E: Logistics and Transportation Review*, 155.
- Fuentes, G. and Adland, R. (2023). Greenhouse gas mitigation at maritime chokepoints: The case of the panama canal. *Transportation Research Part D: Transport and Environment*, 118:103694.
- G2 Ocean AS (2022). Annual report 2021.
- G2 Ocean AS (2023). About g2 ocean.
- Gkerekos, C., Lazakis, I., and Theotokatos, G. (2019). Machine learning models for predicting ship main engine fuel oil consumption: A comparative study. *Ocean Engineering*, 188:106–282.

- Greff, K. K., Rupesh, R. K., Koutnik, J., Steunebrink, B. R., and Schmidhuber, J. (2017). Lstm: A search space odyssey. *Transactions on Neural Networks and Learning Systems*, 28(10):2222–2232.
- Gui, D., Wang, H., and Yu, M. (2022). Risk assessment of port congestion risk during the covid-19 pandemic. *Journal of Marine Science and Engineering*, 10:150.
- Heij, C. and Knapp, S. (2015). Effects of wind strength and wave height on ship incident risk: regional trends and seasonality. *Transportation Research Part D: Transport and Environment*, 37:29–38.
- Helms, M., Ault, S. V., Mao, G., and Wang, J. (2018). An overview of google brain and its applications. In *Proceedings of the 2018 International Conference on Big Data and Education*, pages 72–75.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- International Maritime Organization (2023). Imo identification number schemes.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Knive, R. and Liane, J. M. (2022). Estimating the impact of weather-sensitive cargo risk on transport cost. Master’s thesis, Norwegian School of Economics.
- Kreinovich, V., Nguyen, H. T., and Ouncharoen, R. (2014). How to estimate forecasting quality: A system-motivated derivation of symmetric mean absolute percentage error (smape) and other similar characteristics. *Departmental Technical Reports*.
- Kuroda, M. and Sugimoto, Y. (2022). Evaluation of ship performance in terms of shipping route and weather condition. *Ocean Engineering*, 254.
- Leachman, R. C. and Jula, P. (2011). Congestion analysis of waterborne, containerized imports from asia to the united states. *Transportation Research Part E: Logistics and Transportation Review*, 47:992–1004.
- Lee, J. M. Y. and Wong, E. Y. C. (2021). Suez canal blockage: an analysis of legal impact, risks and liabilities to the global supply chain. *MATEC web of conferences*, 339(01019).
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Mano, A. (2022). Access to brazil’s paranagua port blocked by landslides - authority. *Reuters*.
- Mao, W., Rychlik, I., Wallin, J., and Storhaug, G. (2016). Statistical models for the speed prediction of a container ship. *Ocean engineering*, 126:152–162.
- Marine Traffic (2023). Analytics for ports and terminals.
- Martinelli, L. A., Batistella, M., Silva, R. F. B. d., and Moran, E. (2017). Soy expansion and socioeconomic development in municipalities of brazil. *Land*, 6(3):62.
- Meersman, H., Van de Voorde, E., and Vanelslander, T. (2012). Port congestion and implications to maritime logistics. *Maritime Logistics*, pages 49–68.

- Mohanty, S., Pozdnukhov, A., and Cassidy, M. (2020). Region-wide congestion prediction and control using deep learning. *Transportation Research Part C: Emerging Technologies*, 116(102624).
- NASDAQ (2023). About nasdaq.
- Nilsson, J. and Nilsson, M. (2021). Estimating weather margin seasonality in shipping using machine learning. Master's thesis, Norwegian School of Economics.
- Notteboom, T., Pallis, T., and Rodrigue, J.-P. (2021). Disruptions and resilience in global container shipping and ports: the covid-19 pandemic versus the 2008–2009 financial crisis. *Maritime Economics & Logistics*, 23:179–210.
- Notteboom, T. E. (2006). The time factor in liner shipping services. *Maritime Economics & Logistics*, 8(1):19–39.
- Notteboom, T. E. and Vernimmen, B. (2009). The effect of high fuel costs on liner service configuration in container shipping. *Journal of transport geography*, 17(5):325–337.
- Olah, C. (2015). Understanding lstm networks.
- O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Kerastuner. <https://github.com/keras-team/keras-tuner>.
- Oyatoye, E., Adebisi, S. O., Okoye, J., and Amole, B. (2011). Application of queueing theory to port congestion problem in nigeria. *European Journal of Business and Management*, 3(8):24–36.
- Pang, K. (2020). Multi-step multivariate time-series forecasting using lstm.
- Park, K., Sim, S., and Bae, H. (2021). Vessel estimated time of arrival prediction system based on a path-finding algorithm. *Maritime Transport Research*, 2.
- Peng, W., Bai, X., Yang, D., Yuen, K. F., and Wu, J. (2022). A deep learning approach for port congestion estimation and prediction. *Maritime Policy & Management*, pages 1–26.
- Portner, H.-O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegria, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and (eds.), N. W. (2019). Summary for policymakers. In *In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, IPCC.
- Portos do Paraná, Port Authority (2023). General information.
- Pruyn, J., Kana, A., and Groeneveld, W. (2020). Analysis of port waiting time due to congestion by applying markov chain analysis. In *Maritime Supply Chains*, pages 69–94. Elsevier.
- Qiu, Y., Zhou, J., Khandelwal, M., Yang, H., Yang, P., and Li, C. (2021). Performance evaluation of hybrid woa-xgboost, gwo-xgboost and bo-xgboost models to predict blast-induced ground vibration. *Engineering with Computers*, pages 1–18.
- Shin, D. H., Chung, K., and Park, R. C. (2020). Prediction of traffic congestion based on lstm through correction of missing temporal and spatial data. *IEEE Access*, 8:150784–150796.

- Shipping and Freight Resource (2023). What is liner shipping and what is its role in container shipping..??
- Sirimanne, S. N., Hoffman, J., Juan, W., Asariotis, R., Assaf, M., Ayala, G., and Premti, A. (2019). Port congestion and implications to maritime logistics. *Review of maritime transport 2019. United Nations conference on trade and development*.
- Statista (2023). Leading soybean producing countries worldwide from 2012/13 to 2022/23.
- Steven, A. B. and Corsi, T. M. (2012). Choosing a port: An analysis of containerized imports into the us. *Transportation Research Part E: Logistics and Transportation Review*, 48:881–895.
- Swalin, A. (2007). Choosing the right metric for evaluating machine learning models — part 1.
- Taskar, B. and Andersen, P. (2020). Benefit of speed reduction for ships in different weather conditions. *Transportation Research Part D: Transport and Environment*, 85.
- UN Statistics (2020). Overview of ais dataset.
- U.S Department of Agriculture (2023). Brazil’s momentum as a global agricultural supplier faces headwinds.
- Wang, K., Yan, X., Yuan, Y., and Li, F. (2016). Real-time optimization of ship energy efficiency based on the prediction technology of working condition. *Transportation Research Part D: Transport and Environment*, 46:81–93.
- Ward, R. (2016). New private port planned for brazil’s paranagua.
- Watson, R., Holm, H., and Lind, M. (2015). Green steaming: A methodology for estimating carbon emissions avoided. In *Thirty Sixth International Conference on Information Systems, Fort Worth 2015*.
- Weerakody, P. B., Wong, K. W., Wang, G., and Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, 441:161–178.
- Wijaya, M. (2023). Port of rotterdam’s bunker sales up 5.9% in 2022.
- Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82.
- World Port Source (2023). Port of paranaguá.
- Wu, Y., Tan, H., Qin, L., Ran, B., and Jiang, Z. (2020). A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C: Emerging Technologies*, 90:166–180.
- Yang, D., Wu, L., Wang, S., Jia, H., and Li, K. X. (2019). How big data enriches maritime research—a critical review of automatic identification system (ais) data applications. *Transport Reviews*, 39(6):755–773.
- Yeo, G. T., Roe, M., and Soak., S.-M. (2007). Evaluation of the marine traffic congestion of

- north harbor in busan port. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 133:87–93.
- Zheng, J., Zhang, H., Yin, L., Liang, Y., B., W., Li, Z., Song, X., and Zhang, Y. (2019). A voyage with minimal fuel consumption for cruise ships. *Journal of Cleaner Production*, 215:144–153.
- Zis, T. P., Psaraftis, H. N., and Ding, L. (2020). Ship weather routing: A taxonomy and survey. *Ocean Engineering*, 213.

Appendix

A1 Methodology Illustration

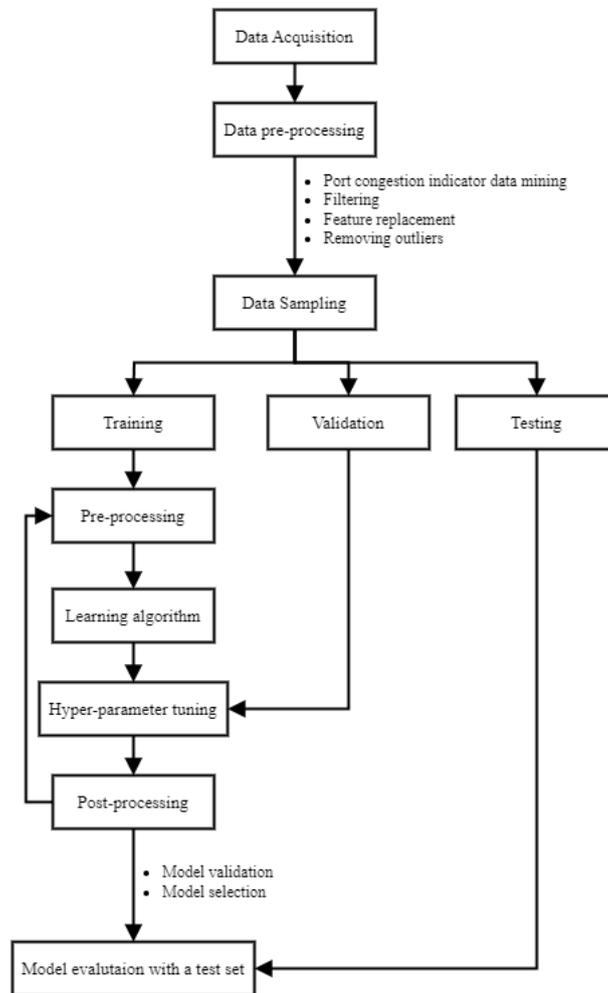


Figure A1.1: Methodology inspired by Abebe et al. (2020)