Vessel Activity Prediction. Final Report - Group Robin1

Joachim, Mick, Wessel, Wesley, and Danila

January 29th, 2021

1 Introduction

1.1 Problem Description

A major problem in maritime service is predicting what a ship's next destination will be and when it will arrive. Despite the ship's captain having knowledge about the next port destination and the expected associated travel time, this is often not logged and communicated extensively with all the maritime and port-maintenance providers. As a result, maritime and maintenance industries can plan their business processes only to a limited extent. The most important piece of information that is needed to improve the scheduling practices of vessel maintenance is knowing which port will be the next destination of the ship. Furthermore, the window of time where maintenance can take place is dependent on the arrival and exit times at that next destination.

Predicting a ship's next port destination is a difficult problem. There are thousands of active ports in the world, all of which have the potential to be the next destination for a vessel. Of course there are certain ports that are not visited often, or that are so unlikely to be the next visit of a ship that they can basically be eliminated as a possible next port destination. Yet, even by logically eliminating the extremely unlikely candidates, many potential next port candidates remain. The travel duration and especially the stay duration in the next port are dependent on many different factors. For example, the travel duration depends on the nautical distance between the ports, the speed of the ship and the weather. The stay duration is likely to depend on the ship type, the size, whether the ship is loading or unloading its cargo, and the weather. These are just examples, in reality there are many more factors that influence the next port, the travel duration, and the stay duration.

1.2 Objectives

Based on the problem description there are three distinct objectives; to predict (1) the next destination port of a vessel, (2) the travel duration from the current port to the next destination, and (3) the stay duration at the destination port. If these three objectives are properly predicted, the maintenance crew can be sent to the predicted port to provide maintenance during the time the vessel is at that port. This also means that if the predicted stay duration is shorter than the expected time it takes to provide maintenance, the scheduling operator can decide to forgo maintenance at that specific port and plan it for the next destination.

1.3 Approach

To be able to predict the three aforementioned objectives, data is needed. Fortunately, a lot of data is available. For most big vessels, historical mooring locations and times are known. The data is produced by transmitters that are unique to each vessel. These transmitters broadcast their current location, which can be matched to specific ports. However, oftentimes the data is not of high quality due to noise, processing errors, and other external influences. Because of the availability of data and the rise of Machine learning, it should be possible to produce predictions of the next destination and journey time of ships at sea. In this project - which is hosted by the company Pon - travel time, stay duration and port destination of vessels at sea will be predicted using machine learning algorithms. For each objective a model will be developed. Then, these models will be combined to make a full prediction of the relevant information. The predictions will be comprehensively shown in a Decision Support System, which can be used to obtain predictions to improve maintenance scheduling.The Decision Support System allows a user to retrieve the predictions for a specific vessel, but also to get an aggregate prediction for a specific port. For example, the expected number of vessels that will visit the port in the coming week. This specific information is especially valuable for harbour masters. In this report the final Decision Support System (User Manual), as well as the underlying machine learning models (Scientific Explanation, Experimental Setup, and Verfications) will be described. Finally, a conclusion will be formulated about the performance, usability, and implications of the current project. Further documentation of the DSS and the code is provided in a separate document.

1.4 Relevant Work

Predicting ship routes, with the corresponding characteristics, such as travel time and stay duration, is a much sought-after problem. Data availability has increased since the implementation of the Automatic Identification System Transponders (AIS Transponders) for large ships. One of the problems many data scientists encounter is that the available data also contains many inconsistencies (e.g. incorrect or missing data is present). Scientists at the Naval Academy of Bulgaria already investigated the problems with the available AIS data and how the accuracy and reliability of this data could be improved [1]. They found that merely 77% of the AIS transmitted data is correct. Erroneous data consisted of errors in Call Sign, AIS Type, Antenna Position, Draught, Destination and ETA, Navigational Status and more. The researches argued that the cause of these errors is due to errors made by technicians that install the devices and errors made by crew members that entered the data in the system. The manners in which the data reliability could be improved were not of much interest for the current project, since the solutions concerned training technicians, crew members, and users of the data.

Moreover, research has also been conducted with regard to predicting the destinations of ships. For example, researchers at the Technical University of Dresden investigated how they could make predictions about the next destination and travel duration of a ship in real time [2]. The researchers used various real-time signals, such as GPS, draught, course, and heading, to estimate the live location and to predict the travel times and port destinations. Because they were able to follow the ships in real time, they were able to predict very accurately (97% accuracy rate) at which port the ship would dock. Furthermore, the researches had a limited data set, of just 300 unique vessels. This study aims to solve a similar problem. However, the goal of the current project is to develop a more generic model, where real time information is not needed. While the final performance of this project will not be comparable with the performance achieved by Lam and colleagues, it does give an indication of what is possible with real time data.

Another study, conducted by scientists from the Centre for Marine Technology and Ocean Engineering (CENTEC), focused on the short-term trajectory prediction of ships [3]. They used mathematical models to predict the longitudinal and latitudinal positions of vessels. With the help of these predictions target-tracking, collision avoidance, and identifying abnormal traffic duration can be improved. While these aspects are not necessarily of interest for the current project, they do provide potentially useful characteristics for the objectives of this project. Furthermore, the research by the CENTEC scientiscs, is also based on AIS data and had similar information about the ships.

Another research, conducted by Alessandrini et al [4], focused on the estimated time of arrival of a vessel using AIS vessel data. The paper focuses on path finding for vessels, and combining the found path with the speed of a vessel to predict the travel duration. The path finding is an interesting concept that could be useful for the current project to estimate the nautical distance between ports. Unfortunately the models that were used to compute the optimal path were not published and it was not deemed efficient to try to replicate the models that were build. Such a model would not be necessary if the nautical distance between ports was publicly available. However, since this is not the case, the distance between ports will have to be estimated in the current project.

2 Scientific Explanation

In this section the steps that were taken to prepare for training the models will be explained. First, information will be given about the data that was used and how the preprocessing was done. Then a section will be dedicated to new features that were engineered to improve the performance of the models. The last section is about which models and baselines were selected for each of the three objectives.

Note. The selected models are non-temporal, therefore removing instances during the preprocessing should not affect model performance.

2.1 Data Exploration

Three separate data sets were provided and used in this project; (1) a data set containing visit information, (2) a data set containing vessel information, and (3) a data set containing port information. The (1) visit data set contained information about which vessel entered and exited which port at which time. From this data set the targets for the models could be derived, since for each unique vessel the next port, the time between exiting the current port and entering the next port, and the time between entering the next port and exiting the next port can be computed. Furthermore, information was present about the distance between the moor location of the vessel and the center of the port as well as some other possibly useful features. In total the data set contained $\tilde{2}.75$ million instances. The entry dates ranged from the 2nd of April 2019 to 22nd of December 2020. The (2) vessel data set contained information such as type, speed, length, and depth of a vessel. There was information about almost 5.000 vessels in the vessel data set. However for a part of the vessels either the length or the depth was missing. With respect to the (3) port data set, information was given about the coordinates, the country the port was located in, the name of the port, as well as some other features that did not provide much information. There was information about 3.600 ports in the port data set. The three different data sets can be combined with each other based on vessel ID. (MMSI) and port ID (arbitrary).

After the three data sets were merged some observations were made. For example, in the visit data set more than 70.000 unique vessels were present, while the vessel data set only contained information about 5.000 vessels. This is concerning, since information about the vessels would be of lesser value than previously anticipated. Basically, 90% of the visits would not have information about the vessel type, speed, and size. Contrarily, only around 1.800 ports from the ports data set were visited at least once in the visit data set. This means that only a part of the ports data set will be used, which is not an issue.

Furthermore, some weird patterns were found in the visits data set. For example, sometimes vessels left a port and then returned to the same port without an intermediate stop. The assumption was made that if the time that the vessel left the port was small enough (e.g. less than a couple of hours) it was because the vessel needed to make space for another vessel to enter the port. If the time was larger then it could be an erroneous measurement, or a fishing vessel that went out to fish. No alternate explanations were found as to why a vessel would leave a port for a long time and then return to that same port without visiting another port.

Lastly, some irregularities were found concerning the target variables; travel duration and stay duration. With respect to the travel duration some vessels took a extremely long time to reach the next port (e.g. going from Amsterdam to Rotterdam in three weeks) or reached the next port by traveling with the 'speed of light' (e.g. going from Madrid to Tokyo in under an hour). The stay duration also had some irregularities. For example, vessels that stayed at a port for more than a month or vessels that did not stay at a port at all (e.g. stay duration of 0 minutes). Moreover, both travel duration and stay duration were exponentially distributed, which might become problematic for predicting these values.

Other small irregularities were found, but those are not worth mentioning here. All adaptions to the original data can be found in the final code of the project.

2.2 Preprocessing

The irregularities found during the data exploration phase needed to be solved. This is done during the preprocessing phase. Data preprocessing is crucial in any data related process. Especially Machine Learning related methods are directly impacted by the quality of data preprocessing. Data preprocessing has the ability to reduce the complexity of the data, and to remove erroneous data or noise from the data. The first preprocessing step that was taken was to select only the useful columns in each data set. For example, the columns containing the maximum depth and length of a port had more than 95% of the values missing. The values were only present for some ports in Madagascar. Next, the missing values of depth and length in the vessel data set were filled. This was done by a simple linear regression that predicted the length/depth based on the other features. This was a highly accurate model since the length and depth were highly correlated and almost all instances had at least one of them present. However, as stated in the data exploration only 10% of the instances in the visit data set could be matched to a vessel in the vessel data set. Therefore, the decision was made to categorize the length and depth of a vessel and fill the missing values in the visits data set with an 'Unknown' label. The chosen categories were based on the 25th, 50th and 75th quantile values.

The second irregularity, weird patterns in the visits data set, was solved by combining visits that exited a port and then entered the same port in a small time frame. For example, A vessel which had two visits, one with an original entry time of 9am and an exit time of 11am, and a subsequent entry at the same port at 1pm and left it again at 4pm to go to a different port, were combined to a single new visit where the entry time is 9am and the exit time is 4pm. This also means that the new stay duration would become 7 hours. If the time outside the port was longer than 12 hours the visits would remain separate.

The last irregularities group concerned the travel duration and stay duration. On average ships travel at 20 knots per hour, this is equal to about 37 km/h [6]. Based on the coordinates of the ports an estimate of the distance between ports can be made. Note that this estimate does **not** account for distance by water/sea, so the estimate is an absolute lower bound estimate. Any visit where the estimated lower bound of speed was larger than 50 km/h was removed. Visits where the estimated speed was extremely low (i.e. less than 0.1 km/h) were also removed from the data set, but only if the next port differed from the original port. Furthermore, visits with a extremely large travel duration were removed. The largest sea-distance is between Finland and Korea and is about 23.711 km [7]. With the average speed of a vessel of 37 km/h one should be able to reach Korea in 640 hours. To make this limit a bit more lenient for slower vessels [5] (i.e. those who travel at 12 knots per hour) it was decided to remove any visits with a travel duration of more than 1.250 hours. Visits with a very short travel duration (i.e. less than 6 minutes) were also removed. With respect to the stay duration, the visits where the stay duration was 0 minutes were removed. The visit before and after the removed visit were connected by changing the target port in the first instance and the current port in the second instance. Since there is no logical reasoning to decide a explicit cut-off point for visits with a very long stay duration, it was decided to trim the 1% of visits with the longest stay duration. This should remove the most influential outliers. Lastly as explained in the data exploration the travel duration and stay duration were exponentially distributed. Therefore, the natural logarithm was taken of these variables. This resulted in relatively normally distributed target variables.

Table 1 describes the most impactful actions taken in the preprocessing phase and the number of instances that were affected. The last two points have not been discussed yet. An assumption that was made for this project was that when a user would look up a vessel, that specific vessel should already be present in the visits data set. After some inspection it was found that there were singular instances, which means that no previous port or target port were present for an instance. Only the current port was known. Since these instances violate the assumption that was made, it was decided to remove these singular instances. With respect to the last point; the vessel ID code (i.e. MMSI) is a 9-digit unique identifier for a vessel. There were some vessel ids present in the data set with less than 9 digits, so the assumption was made that these unique identifiers were erroneous and it was decided to remove these few instances. Furthermore, based on the MMSI the Maritime Identification Digits can be extracted. These denote for example the country of origin. However, the MIDs also provide information whether the object is of a vessel group or a search and rescue vehicle (e.g. helicopter). So the MMSIs with an MID that indicate that the object is not a 'regular' vessel were removed. Note that the order of the actions in the table was also the order in the code. This could for example mean that there might have been more erroneous MMSIs that were already removed in previous steps. At the end the data set contained just over 2 million instances (originally 2.75 million).

Step	Action	Affected Instances
1.	Stay Duration not 0 minutes	213.315
2.	Combine Visits same port	216.673
3.	Speed not faster than 50 km/h	46.721
4.	Speed not slower than 0.1 km/h	45.382
5.	Travel Duration max 1.250 hours	55.017
6.	Travel Duration min 0.1 hour	602
7.	Stay Duration trim 99%	20.869
8.	Singular instances	16.216
9.	Remove invalid vessel ids	44

Table 1: Preprocessing Steps

2.3 Feature Engineering

After preprocessing the data the next step is feature engineering. Feature engineering is an important part of any Machine Learning project, since it can drastically boost the performance of the used models. Based on literature, data exploration, and common sense several topics for feature engineering were identified;

1. Port History
2. Previous Port Information
3. Distance Estimates
4. Ship Information
5. Frequent Connections
6. Speed and Travel Time Estimates
7. Comparison Features

This information will be helpful to predict the next port destination, travel time and stay duration at the destination of a vessel. All features will be included in the models for each of the three objectives. The models themselves will be able to deduce which features are especially relevant for each objective. To be able to compute some of these features, historic data is needed. Any leakage of the historic data into the training and or test set would heavily reduce the reliability and validity of the final product. Therefore, the oldest 50% of the data (i.e. about 1 million instances) was split of from the data set. The oldest data is selected based on the entry date of the visits. The actual split date was the 31st of January 2020.

Features that would be included for (1) Port History are for example the port popularity (i.e. number of visits), the type of vessels that have visited that port, and the original features from the port data set. These features are then treated as constants that describe each port. Therefore, they can be reused to provide information about the (2) Previous Ports. Based on the country codes and coordinates of each port, the (3) distance between ports can be estimated. The nautical distance between countries was retrieved using the CERDI data set [7]. This can be an indication of how far a ship tends to travel. Note that the nautical distance is 0 if two ports are located in the same country. Another distance metric is the distance between ports, which can be estimated based on the port coordinates. While this distance metric does not account for paths that ships can travel (i.e. ships cannot go over land), it could potentially still be useful to estimate the travel times.

(4) Ship Information is about features based on the historic travels of a vessel as well as the original variables from the vessel data set. Examples of such features would be the number of (unique) ports a vessel has visited, the average travel duration, and the average stay duration. It is likely that not all vessels have enough historic data to compute accurate values for these features, but they could still be useful for those vessels that do have enough data. Furthermore, the MMSI of a vessel can be used to derive the country of origin (MID) and the region of origin, which could be valuable information. (5) Frequent Connections can also be derived from the historic visits. Features based on this information include the average travel time from one port to another, and the frequency that vessels go from one port to another. (6) Based on the estimated distance between two ports and the historic travel times, the actual travel time to another port can be estimated. For example, if a vessel travels at an absolute lower bound speed of 16 km/h and the absolute lower bound of the distance between two ports is 160 km, the estimated travel time would be 10 hours.

Lastly, (7) the previous port information, the current port information, and the target port information can be compared. For instance, 40% of the vessels that go to the current port are tankers and 0% of the vessels that go to the next port are tankers. A simple prediction would be that the vessel will not visit the next port. Obviously this information could not be used in a normal classification model since the target is not known. Therefore, as will be explained in the model selection a ranking model was chosen where several ports are ranked on likelihood that it will be the next port. Altogether, these topics could provide information that can boost the performance of each model that will be trained. At the end of the feature engineering phase there were around 300 features.

2.4 Model Selection

As stated in the introduction, the current project has three objectives: to predict the next destination, to predict the travel duration, and to predict the stay duration in the next port. To achieve the first objective a classifier of some sorts is needed, while for the other two objectives a regression model would suffice. Since the data is of temporal nature a temporal model could be selected. However, due to the previously mentioned irregularities the expected performance of a temporal model was low. Therefore, it was decided to use non-temporal models for each objective. Furthermore, since a lot of missing data was still present after the preprocessing, the Light Gradient Boosting Machine (LightGBM) library was chosen for the models. The LightGBM library is a gradient boosting framework that utilizes tree-based learning algorithms and is able to handle missing data. To be able to compare the performance of the models a baseline is also needed. The exact models and corresponding baselines will be described for each objective below.

2.4.1 Port Prediction

There are 1.800+ ports that have been visited all across the world based on the historic data. This means that the selected model has to be suited to handle many classes. As stated it was decided to use a non-temporal model. While basic classifiers will probably have difficulties handling a large number of classes, a ranking model would not have this problem. A ranking model (LightGBM Ranker) expands a single instance into multiple instances, each with a different target class. Then the model ranks these multiple instances based on the likelihood that it is the correct class. By using this type of model it is possible to use all of the previously engineered features. For example, the number of connections between the current port and each possible class (target port) can be added as a feature. In normal classifiers this would not be possible since the target class is unknown. Of course the target variables from the second and third objective were not used as predictive features in this model.

An issue that arises from using a ranking model is that it is nearly impossible to add all possible target classes and rank them. This would mean that each of the 1 million training instances have to be expanded to 1.800+ ports leading to a data set with 1.800*1 million instances. To moderate this a random selection of training instances will be made as well as a selection of possible target ports. The selection of possible targets should not be done randomly since most ports in the port data set are very unlikely to be the target port, leading to a very easy objective for the model. Therefore, the 25 most common historic connections from the current port as well as 75 random ports were used as possible target ports. This results in a set of possible target ports where some are likely to be the next port, while most others are unlikely. This situation corresponds to the final Decision Support System where all 1.800+ ports will be ranked on their likelihood to become the next port. The baseline that was used to compare the model performance is a ranking based on the number of historic connections (e.g. a vessel leaving Amsterdam is most likely to go to Rotterdam). The historic connection baseline will be provided with the historic data set that was split of from the rest of the data as well as the complete training data. The performance of both the model and the baselines will be evaluated based on the test set.

2.4.2 Travel Duration & Stay Duration

As opposed to predicting the next destination, predicting the travel duration and stay duration are regression problems. The model that was selected for both objectives was the LightGBM Regressor. All the engineered features were used. Obviously the target variable for stay duration was not a predictive feature for predicting the travel duration. For these two models no special operations have to be taken. The baseline that will be used for both of these models is the average travel duration and average stay duration when traveling from the current port to the target port. For example, if there are 200 visits that go from Amsterdam to Rotterdam, the baseline will predict the average travel duration and the average stay duration when in Rotterdam of those 200 visits.

3 Experimental Setup

3.1 Train test Split

After the models for each objective have been selected they have to be trained. The preprocessed data set with the new engineered features will be split up into three parts; the train set, the validation set, and the test set. The models will be trained on the train set, after which the validation set will be used to initially check for the performance. Based on the performance on the validation set, the hyperparameters of each model will be tuned. After achieving the desired performance on the validation set the models will be evaluated on the test set. It is important to make sure that the train, validation, and test splits are performed in a reasonable manner. This means that, since the data is of temporal nature, the train set should contain the oldest data, while the test set should contain the most recent data. Based on standard Machine Learning practices it was decided to split the data into 60% train set, 20% validation set, and 20% test set.

3.2 Evaluation Metrics

To be able to gain knowledge on the performance of a model (or the baseline) evaluation metrics are needed. Besides being able to compare the final performance, evaluation metrics are also important for hyperparameter tuning and feature selection. Since there are two types of models (ranking and regression) that will be used in this project the evaluation metrics will be discussed separately for each type of model. An overview of the used metrics is given in table 2.

3.2.1 Port Prediction

In order to distinguish the different qualities of various models, the decision was made to use multiple metrics, which together determine the predictive power of a model. After all, one model can, for example, have a relatively high accuracy score (the model predicts the vessels' next port correctly the most), yet when it does not predict the next port correctly, it is extremely wrong (e.g. the model predicts that the ship in question will travel to Northern-America, when in fact it is going to Africa) Obviously, this should affect the general predictive score that is assigned to the model. That is why it is important not to look purely at a single metric, but to take multiple metrics which give different insights in the nature of the given predictions into account. This consideration has led to the development and implementation of several metrics. The accuracy and f1 score indicate the degree of precision of the model, to what extent the model can correctly predict the next port. The Topx metric provides insight into the predictive qualities of the model, also beyond the prediction of the most likely port. Moreover, it offers a way to gain insight into the type of prediction or in which area the model expects the ship to travel to. Lastly, Port distance is a computation of the distance in Kilometers between the predicted next port and the actual next port, so we can see if the model predictions are near the actual next port destination.

3.2.2 Travel Duration & Stay Duration

Predicting the travel and stay duration of a particular ship is of a similar nature. In both cases, a numerical value must be predicted on the basis of (historical) data from the ship. The high variation of both the stay duration and the travel duration makes this issue a complicated problem. To get insight into the performance as well as the areas where the models perform worse, the use of multiple metrics is preferred. Therefore, metrics are used that describe the mathematical differences, in absolute terms (RMSE) or in relative terms (MAPE) between the model predictions and the real observations. But also interpretable metrics, which deal with the predictive power of the model in a practical sense (WithinX and Rsquared) are used to evaluate the models. Since the travel and stay duration are in $\mathbb{R}^+$ the X of WithinX has to be chosen accordingly. For example if $X = 100\%$ the score would be optimal if the model always predicts 6 minutes (the lowest travel duration). Therefore it was decided to use Within50% as a metric, which reflects the percentage of predictions that are within 50% of the target value. Together, they provide a basis for the predictive

Objective	Metric	Description
1	F1 score	*Harmonic difference of precision(% of true positives) and recall (% of positive classified instances.)*
1	TopX	*The fraction of times the correct prediction was within the top X predictions.*
1	Accuracy	*The fraction of predictions that the model predicted correctly.*
1	RX-Accuracy	*The fraction of predictions that are within a radius of X kilometres from the target port.*
2/3	(R)MSE	*A mathematical way of calculating the total error between a prediction and the actual value.*
2/3	MAPE	*The mean absolute percentage error describes the average deviation in percentage between the predictions and the actual values.*
2/3	WithinX	*The fraction of times the prediction was within x% of the real value.*
2/3	Rsquared	*The (fraction of) variability explained by the statistical model.*

Table 2: Metrics used for model evaluation

power of each model, so that, given the pros and cons of each, the model that offers the best solution in terms of predictions can be chosen.

3.3 Feature Selection

Now that the data is split and the evaluation metrics are defined, it is possible to move on to feature selection. As stated in Feature Engineering there were around 300 features present after adding the newly engineered features. While this number is not that large in comparison with other studies it was decided to still use a form of feature selection. This decision was made, because feature selection increases the computational efficiency as well as training speed while only slightly decreasing the performance. Each previously specified model was trained on the same set of features with 500 estimators. Next, the feature importances (split-based) were extracted from the model. The 50 most important features were selected and used in the final model. Note that the selected features were not shared across the models. For example, the historic travel duration between ports was an important feature for predicting the travel duration, while it was not a relevant feature for the port selection model. Based on some preliminary analysis of the performance on the validation set it was concluded that the model performance did not significantly drop after selecting just 50 features. More features would have been selected if this was the case.

3.4 Hyperparameter Tuning

Since all models used were from the LightGBM library and were thus tree-based, all of the models share the same hyperparameters. For each objective, models were trained with varying hyperparameters on the train set and then evaluated on the validation set. The

Parameter	Port Prediction	Travel Duration	Stay Duration
metric	NDCG@1	MAPE/RMSE	MAPE/RMSE
n_estimators	2.500	5.000	5.000
max_depth	4	15	10
max_num_leaves	10	50	15
learning_rate	0.025	0.025	0.05
early_stopping	250	250	250
min_child_samples	25.000	500	5.000

Table 3: Hyperparameter Settings

hyperparameters were tuned manually, so no functions such as grid search were used. This was done due to the fact that training a model on over 1 million instances takes a considerable amount of time, so grid search was not deemed an efficient method. The goal of the hyperparameter tuning was to maximize the performance on the validation set while minimizing any overfitting on the train set. The final selected hyperparameters are shown in table 3.

4 Verifications

4.1 Results

4.1.1 Port Prediction

The LightGBM Ranker was trained and evaluated according to the previously described experimental setup. It is important to reiterate that the model was trained and initially evaluated (validation set) on ranking 75 ports of which some were randomly selected and others were selected based on frequent connections. The results described in this section were obtained from the instances in the test set where for each instance all 1.800+ ports were ranked. This means that there is a reality gap between the performance on the validation set and on the test set. However, the performance on the validation set will not be discussed. Because the model is a ranking model, the accuracy would be reflected by the percentage of correct predictions of the highest ranked port prediction. The model achieved an accuracy of 38.4%, in comparison the baseline model achieved an accuracy of 27.5%. To further inspect the performance of the rankings of the model and the baseline the TopX measure was plotted in figure 1. Note that the TopX metric with $X = 1$ is the same as the accuracy score. Figure 1 shows that the TopX score of the model is higher than that of the baseline for the first few ranks. However, the performance of the baseline seems to be higher for the ranks 15-60. After that the model performance becomes higher than the baseline again. While the performance of the baseline is higher for some TopX it can be said that the model is more useful. This is because when predicting the next port it doesn't matter what the 15th rank is that is predicted. The most important ranks are the first few, since those are the most likely destinations. The next metric is RX-Accuracy which is plotted in figure 2. Figure 2 shows that the model reaches an RX-Accuracy of 54.2% when taking a radius of 100 kilometres around the target port. This means that, while the prediction may not be correct, half of the time it is only an 1 hour drive away from the actual port. Figure 2

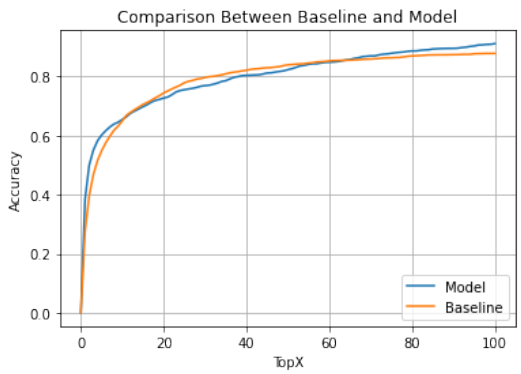

Figure 1: TopX

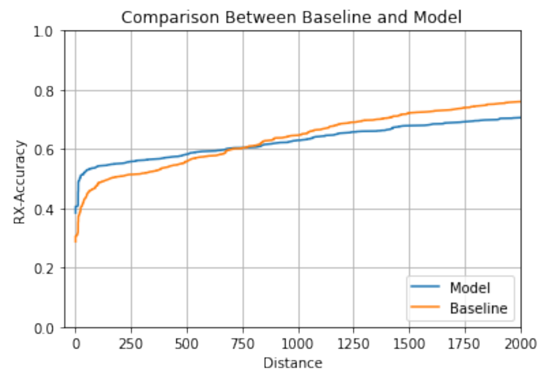

Figure 2: RX-Accuracy

also shows that the performance of the model is better than that of the baseline. However, for larger radii the performance of the baseline is better, but larger radii are not of any relevance within the current application context. Note that the graphs for the baseline and the model start at the accuracy scores 38.4% and 27.5% respectively. This is because the distance in those cases is 0 kilometres. Lastly, the f1 scores were computed. The f1 score of the model is 0.127, while the f1 score of the baseline is 0.087.

4.1.2 Travel Duration

The model for predicting the travel duration between ports was trained following the experimental setup. The perfomance on the RMSE, MAPE, Within50% and R^2 of the LightGBM Regressor are schematically shown in table 4. Table 4 shows that the performance of the model is better than the baseline on every metric. The actual performance might seem relatively low, however, it must be noted that the travel duration ranges from 15 minutes to 1.250 hours. This means that the MAPE explodes when the prediction is slightly off for short travel durations, while on the other hand the RMSE explodes when the prediction is relatively not far off for very long travel durations. To visualize this the RMSE and MAPE measures were plotted against the travel duration. This is shown in figure 3. Figure 3 shows that the performance of the model is better for all travel durations than the performance of the baseline. Furthermore, it can be seen that the MAPE is indeed very high for short travel durations while the RMSE for those is very low. The same is true for larger travel durations, for those the MAPE is very low while the RMSE becomes very large. For practical purposes of these predictions one should look at the RMSE for short travel durations and at the MAPE for longer travel durations.

metric	model	baseline
RMSE	94.08	117.92
MAPE	76.9%	98.0%
Within50%	53.1%	42.3%
R^2	60.1%	40.0%

Table 4: Results Travel Duration

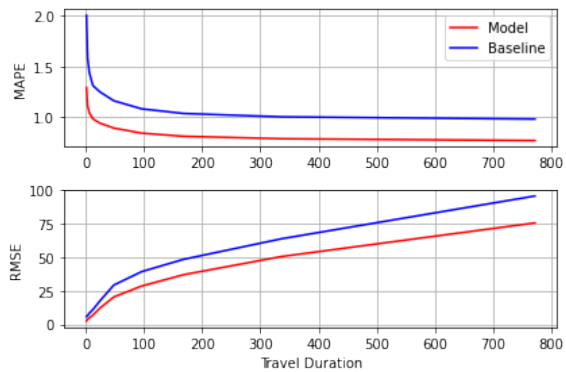

Figure 3: Performance Travel Duration

4.1.3 Stay Duration

The same metrics that were used to evaluate the performance of predicting the travel duration were used to evaluate the performance of predicting the stay duration. The results are shown in table 5. This table shows that, while the performance of both the model and the baseline are quite similar, the model is slightly better on all metrics. The RMSE and MAPE metrics were plotted against the target stay duration to further inspect the found values. Especially on the RMSE a clear difference can be seen between the model and the baseline. The model outperforms the baseline over all possible values for the stay duration. In contrast the baseline and model were very similar on their MAPE metrics over the range of the target variable. An interesting observation is that the RMSE starts very high for both models on the smaller stay durations. This is contrary to the expectation, since the RMSE is usually smaller for the smaller target values. This unexpected result required further inspection. A potential explanation that was found was that due to taking the natural logarithm of the stay duration the target values became normally distributed. This normal distribution had parameters $M = 2.97$ and $SD = 1.41$. This value reflects a stay duration of $e^{2.97} = 19.5$ hours. So it could be that the model is more accurate in its predictions around this value, thus resulting in the RMSE graph of figure 4.

metric	model	baseline
RMSE	101.42	108.49
MAPE	215.1%	238.4%
Within50%	45.9%	40.1%
R^2	14.7%	2.7%

Table 5: Results Stay Duration

4.2 Discussion

The goal of this project was to predict three pieces of information; (1) the next port, (2) the travel duration, and (3) the stay duration. While the predictions for the next port might not

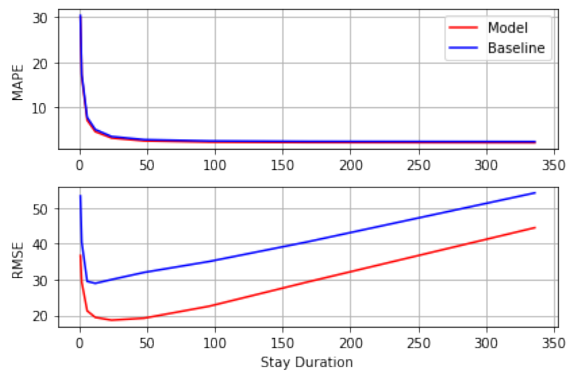

Figure 4: Performance Stay Duration

have been optimal they are still very much usable for scheduling maintenance and harbour masters. Compared to the accuracy that was achieved by Lam et al. [2] the accuracy of the current project seems to be underwhelming. However, as explained in the introduction Lam and colleagues had more relevant information that was not used in the current project. Furthermore, the port prediction model is able to predict the next port for each vessel that has at least two historic connections. Two historic connections is very little and almost all vessels will have that information. The model performance might have been better if more historic connections were used, but this would reduce the generalisability of the model. For example, Discrete Time Markov Chains where the complete historic connections of each vessel is used to predict the next port might achieve better performance. However, these predictions would be great for vessels with a lot of historic connections while performing worse for those with little historic data. Overall it can be concluded that the performance on the first objective is at least acceptable.

With respect to the second and especially the third objective the performance of the machine learning models was decent. When taking the actual travel and stay duration into account the predictions were not that far off. However, the predictions could have been a lot more accurate. Accurate and complete vessel information could have had a massive impact on the performance of these two models, because both travel duration and stay duration is heavily dependent on ship type and size. Unfortunately this information was not available for most vessels in the data set that was used. Moreover, knowledge about other factors could also have an impact on the performance. For example, the time that a vessel waits outside a port to enter is included in the travel duration, which is not optimal. Other information that could be useful is weather information or the seadistance between ports. Neither was used in the current project. Taking these shortcommings into account it can be said that the performance is still decent. The estimates might be far off for some cases, but in general the estimates are in the right direction.

5 Decision Support System

The Decision Support System distinguishes between two different types of users. On the one hand, there are people who are interested in the whereabouts of specific vessels, like maintenance crews for example. On the other hand, there are people who are interested in the condition of specific ports, like harbour masters. This Decision Support System is of use to both groups and combines these two functionalities as two different tabs named 'Find Vessel' and 'Find Port' respectively.

5.1 User Manual

In Figure 5, the screen upon opening the Decision Support System, the 'Find Vessel' tab, is shown. There is an interactive world map with the location of each vessel. Options include zooming in and out, panning around the map and downloading the current view as png. When hovering over a certain vessel, the MMSI and coordinates are displayed. Right above the world map is a dropdown menu where vessels can be either selected from the dropdown, or manually searched for. It must be noted that due to computational limitations only a subset, comprising around five thousands unique vessels, is loaded into the system at once. These vessels have in common that they all departed mid-December 2020.

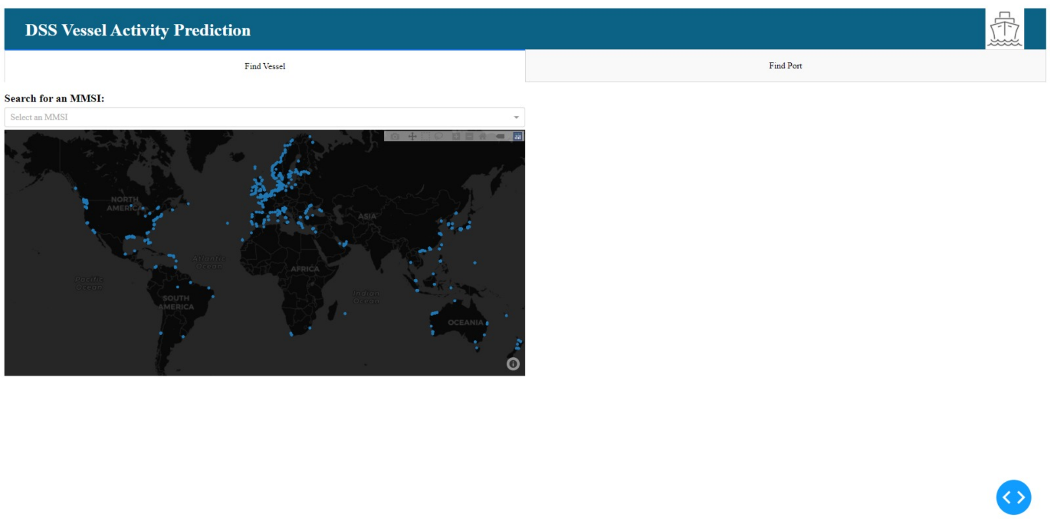

Figure 5: Tab 1: Find Vessel (before selection)

When a vessel is selected in the dropdown menu, the world map updates and two new tables are displayed on the right side of the screen. The world map now only shows the location of the selected vessel (blue), and the location of the predicted next port in a different colour (red). The first table shows some basic information about the vessel like the port and country of departure and its departure time. More information like vessel type and speed could be added to this table, but the large majority of the vessels lack this information. To keep the Decision Support System simple, clean and intuitive it was decided to leave it out. The second table shows what is predicted for this particular vessel. It includes the destination port and country, time of arrival, duration of stay and time of departure. This table brings us to the second tab, 'Find Port', where this information from all vessels is combined to create useful insights.

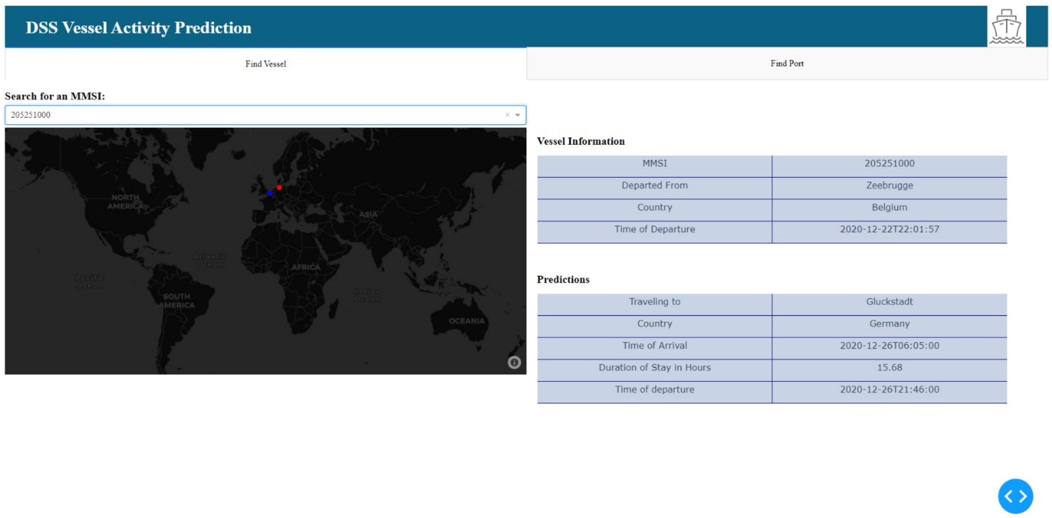

Figure 6: Tab 1: Find Vessel (after selection)

Upon opening the tab 'Find Port', the screen shown in Figure 7 is displayed, regardless of the selection in tab 'Find Vessel'. Similarly, there is an interactive world map with the location of each Port. The options are the same, but the colour of the ports is deliberately different from the vessels to make it clear that tabs are switched. When hovering over a certain port, the port name and coordinates are displayed. Correspondingly, there is also a dropdown menu to select or search for ports. All ports can be selected, also ports without any predicted activity. As port names are not unique, a numerical identifier is added to the search string to distinguish between ports with a similar name.

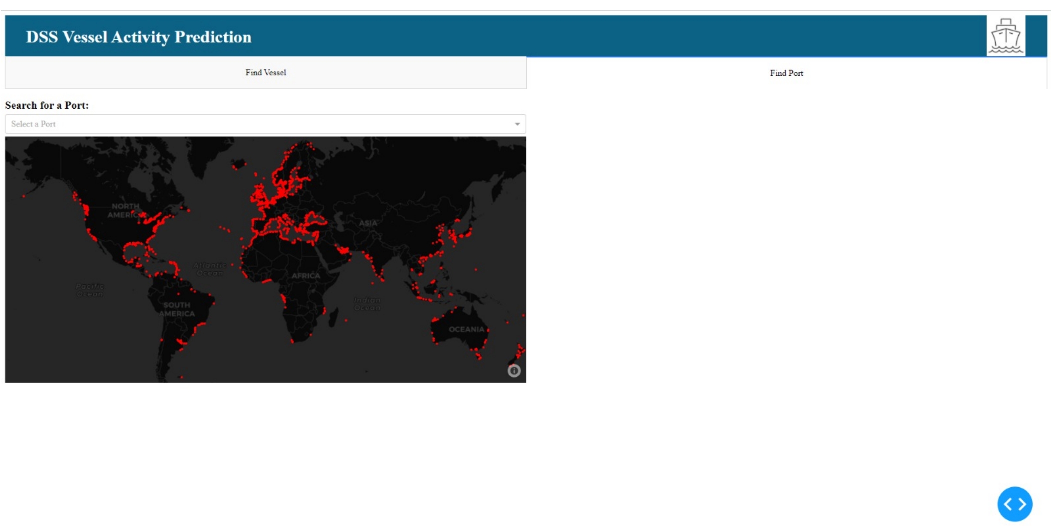

Figure 7: Tab 2: Find Port (before selection)

When a port is selected, three things happen: the world map updates, a table is displayed and a graph is generated. These changes can be found in Figure 8. The world map shows the location of the selected port. The table shows some basic information about location, type and size of the port. Again, it has been decided to leave out information which is missing for most ports. The graph is the most interesting part of the screen. It shows the predicted number of arrivals and departures for this port per day. This could be very useful for harbour masters who want to know how busy it will be in their port on a particular day in the future.

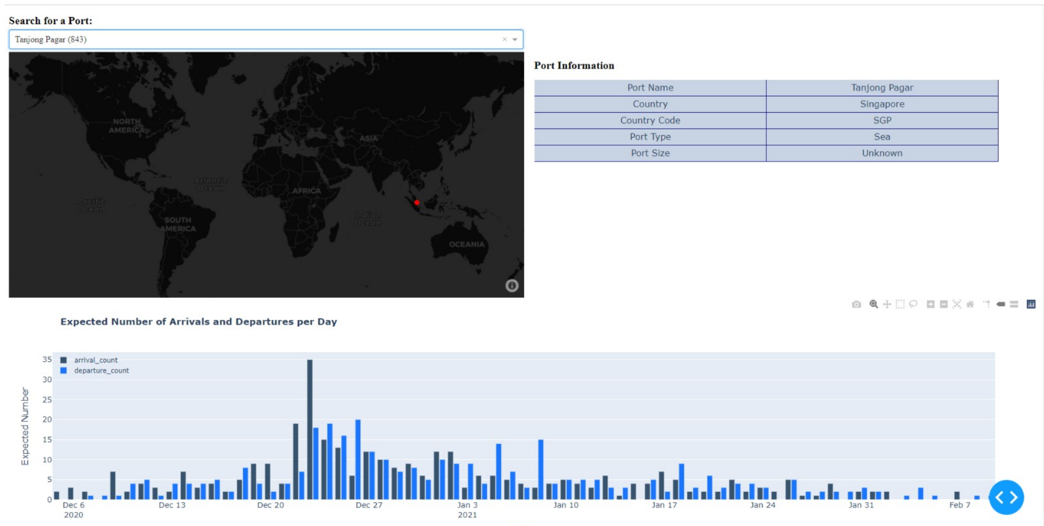

Figure 8: Tab 2: Find Port (after selection)

6 Conclusion

The main objective of the current project was to create a Decision Support System which could be used to access predictions that indicate the next port of a vessel, the travel duration, and the stay duration. This was done by combining AIS data with port information and vessel information. Due to the irregularities found in the data a lot of preprocessing work had to be done. Many features were engineered and three machine learning models were trained and evaluated. The performance of the models was interpreted as good considering the data quality and quantity. However, improvements could definitely made to increase the accuracy of the predictions. The predictions were combined into a Decision Support System. The Decision Support System is an accessible and intuitive application in which one can look up a vessel to retrieve the predictions. Moreover, from the perspective of a port one can use the application to obtain predictions about the number of vessels that will enter and exit a port on a given date. Altogether this system can be used by harbour masters and maintenance planners to prevent errors. Complementary projects should focus on obtaining more relevant, accurate and complete data for each objective. Furthermore, if more time was available the Decision Support System would have been expanded with confidence intervals and probabilities. These were available, but were not included in the application.

References

[1] Sotirov, S., Alexandrov, C. (2017). Improving AIS data reliability. In Global perspectives in MET: Towards Sustainable, Green and Integrated Maritime Transport (pp. 237-244).

[2] Lam, H. T., Diaz-Aviles, E., Pascale, A., Gkoufas, Y., Chen, B. (2018). Grand Challenge: Real-time Destination and ETA Prediction for Maritime Traffic, Arxiv.

[3] Rong, H., Teixeira, A. P., Soares, C. G. (2019). Ship trajectory uncertainty prediction based on a Gaussian Process model. Ocean Engineering, 182, 499-511.

[4] Alessandrini, A., Mazzarella, F., Vespe, M. (2018). Estimated time of arrival using historical vessel tracking data. IEEE Transactions on Intelligent Transportation Systems, 20(1), 7-15.

[5] Vidal, J. (2010). Modern cargo ships slow to the speed of sailing clippers, The Guardian. Retrieved from https://www.theguardian.com/environment/2010/jul/25/slow-ships-cut-greenhouse-emissions

[6] Agarwal, M. (2020). What is the speed of a ship at sea?, Marine Insight. Retrieved from https://www.marineinsight.com/guidelines/speed-of-a-ship-at-sea/.

[7] Bertoli, S., Goujon, M., Santoni, O. (2016). The CERDI-seadistance database, FERDI. Retrieved from https://ferdi.fr/en/indicators/the-cerdi-seadistance-database.

Appendix

A Time Registration

What	Wessel	Wesley	Joachim	Mick	Danila
Meetings	13h	8.5h	13h	10h	2h
Presentations	7h	7h	7h	6h	-
Project Plan	4h	5h	2h	5h	-
Final Report	4h	7h	4h	15h	-
Context Understanding	7h	5h	7h	5h	-
Data Exploration	12h	23h	15h	15h	-
Preprocessing	2h	6h	6h	15h	-
Model Development	3h	20h	3h	30h	-
DSS	28h	-	28h	-	-
Crash Course Dash	-	-	3h	-	-
Docker Configuration	3h	-	-	-	-