# Explainable AI for ship design analysis with AIS and static ship data

Rooij, L. van; Winter, R. de; Kononova, A.V.; Stein, N. van; Vladimir, N.; Malencia, S; Senjanovic. I.

# EXPLAINABLE AI FOR SHIP DESIGN ANALYSIS WITH AIS AND STATIC SHIP DATA

Lucas van Rooij[1,2], Roy de Winter[1,2], Anna V. Kononova[1], Bas van Stein[1]
[1] Leiden University, Leiden Institute of Advanced Computer Science, Leiden, The Netherlands
[2] C-Job Naval Architects, Hoofddorp, The Netherlands

**Abstract.** Decisions made in the early phases of ship design have a large influence on the capital andoperational expenses of a vessel. In order to support decision making in this phase, big data and machinelearning techniques can be of great use. This work shows how Explainable Artificial Intelligence (XAI) and Global Sensitivity Analysis (GSA) combined with Autonomous Identification System (AIS) and static ship data can be used to find important design characteristics of ships. A data collectionframework is setup that collects AIS data over a five month time period. Static ship design data is usedto predict performance related target features that are calculated from AIS data. By applying XAI andGSA methods to the regression models that predict these target features, insight can be gained on howdesign features influence the performance of ships. Experiments showed that for most ship types, theoverall length is the most important design feature for speed related target features. Besides the overalllength, the draught also has a significant impact on the rotation capabilities.

**Keywords:** Explainable Artificial Intelligence, AIS, Machine Learning, Global Sensitivity Analysis.

## 1. Introduction

Over the recent years the application of Artificial Intelligence (AI) in the maritime industry has seen an increase [1]. Examples of recent applications range from fleet allocation to Estimated Time of Arrival (ETA) prediction [2]. These applications have caused a reduction of CO2 emissions and a large increase in efficiency and performance of vessels throughout the whole maritime industry [2]. A subsector of the maritime industry that has only recently seen an increase in applications of AI, is the naval architect industry [3]. Applications of AI and data driven decision making can be of huge benefit for naval architects, as many early phase design decisions previously relied on the experience of engineers or a handful of reference vessels [4, 5]. Any wrong assumptions made in this phase can lead to higher production, or operational costs, which highlights the importance of this phase [6]. In order to come up with more cost efficient and better performing ships, engineers need a clear picture of a ship's desired operational profile. Here, data driven technologies can provide engineers a helping hand. Previous applications of AI systems in the naval architect industry mainly focused on data collected from a relatively small group of vessels. Examples of these are Neural Network aided design of ship hull structures [7], and propulsion power optimum calculation via genetic algorithms [8].

Together with an increase in applications of AI methods, the predictive capabilities of AI models has also increased [9]. As the performance of these models increase, they also tend to become more complex and their workings harder to understand. In order to understand the workings of complex black box predictive models, Global Sensitivity Analysis (GSA) and Explainable AI (XAI) can help. These methods define measures or visualisations of important input features of predictive black box models. In this research these tools are combined with data collected from a large group of vessels in order to obtain a high level view on what ship design parameters are distinctive for a vessel's performance capabilities. The data used in this work consists of Autonomous Identification System (AIS) data, for which a data collection framework is created. Next to AIS data static ship data is provided by C-Job Naval Architects[1]. These data sources will be combined with GSA and XAI techniques in order to quantify the influence of design parameters on the

*Correspondence to: lucas.vr.1997@gmail.com
[1] *C-Job Naval Architects, A sustainable maritime industry in one generation (2022), https://c-job.com/*

speed and rotation related capabilities of ships. The scientific contribution of this work is two-fold: (1) AI and GSA method are applied to a Multi-Output regression model built by an Automated Machine Learning Model. (2) AIS data is used to analyze the world fleet in order to find important design characteristics for commercial ship types. The remaining of this work is structured as follows: Section 2. describes the background, related work and methods used in this work. Hereafter, Section 3. explains how the data is collected, processed, and stored. Section 4. then shows the experiments and results. Finally, Section 5. discusses the results and concludes.

## 2. Related work

In the following subsections related AIS research, relevant background on automated machine learning models, global sensitivity analysis, and explainable AI methods are described.

### 2.1. Related AIS research

The Autonomous Identification System (AIS) has mainly been developed as a system that can be used for collision avoidance and navigational safety for almost all ships in the commercial fleet. The system was initially developed in 2002, but had limited coverage over only the coastal area waters. In 2008 satellite AIS transponders were introduced which increased the coverage to almost all areas in the world. With the increase in coverage and data quality new applications for the AIS appeared. Examples of this are, ship trajectory extraction and prediction [10], ship activity tracking [11], tracking for environmental monitoring [12], or tracking in restricted waters such as locks and canals [13]. Next to vessel tracking, AIS data can also be used to measure various port statistics such as the number of daily visiting ships, or waiting times [14]. To the best of the authors knowledge, AIS data has never been used to compare ship design characteristics of the commercial fleet on a worldwide scale. More on how AIS data is used in this research is explained in Section 2.1.

### 2.2. Automated Machine Learning

Automated machine learning (AutoML) automates the task of applying machine learning techniques to any real world problem. AutoML automates all tasks that one would typically perform manually when designing a machine learning pipeline. By introducing a high level of automation, AutoML allows non-experts to design high performing machine learning models. The most common steps that are performed by AutoML frameworks are: Data preparation, Feature engineering, Model selection, Ensembling, and Hyperparameter tuning all under strict time and memory constraints [15]. An often used implementation of AutoML is the *Autosklearn 2.0* framework [16]. This implementation has gained popularity due to it's high performance on large datasets. The improvements of *Autosklearn 2.0* on it's precedor *Autosklearn 1.0* mainly lie in the model selection strategies, and candidate pipeline selection strategies. For an in depth explanation of the implementation of *Autosklearn 2.0* [16] can be consulted.

### 2.3. Sobol indices

Sobol indices are a form of variance based GSA [17], and can be used to determine the importance of input features. This method decomposes the variance of the output of a machine learning model into fractions that can be attributed to features, or combinations of features. A function $Y = f(\mathbf{X})$ can be decomposed as follows:

$$Y = f_0 + \sum_{i=1}^{d} f_i(X_i) + \sum_{i<j}^{d} f_{ij}(X_i, X_j) + \cdots + f_{1,2,\ldots,d}(X_1, X_2, ..., X_d) \tag{1}$$

With constant $f_0$, and $f_i$ as function of $X_i$. The variance of Y can then be expressed as:

$$V(Y) = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + \cdots + V_{1,\ldots,d}, \tag{2}$$

where

$$V_i = V(E(Y|X_i)), \tag{3}$$

$$V_{ij} = V(E(Y \mid X_i, X_j)) - V_i - V_j \tag{4}$$

Here $E(Y|X_i)$ stands for the expectation over $Y$ given $X_i$, which can be calculated by taking the mean over $Y$ for all $X_i$. The first order Sobol index can then be calculated as a measure of sensitivity $S$ of feature $i$ on model output $Y$ as:

$$S_i = \frac{V_i}{V(Y)}. \tag{5}$$

The first order Sobol index measures the effect of alternating $X_i$ alone averaged over variations of other input features. By dividing it over the total variance it is measured as a fractional contribution. Higher order indices can be calculated by dividing $V_{ij}$, $V_{ijk}$ and so on over $V(Y)$. The first- and higher-order Sobol indices quantify the importance of each feature, or combination of features with respect to the output variance. Evaluating all indices for a large number of features can be problematic as the number of evaluations is quadratic with the number of input features. In practice the total-order Sobol index is often calculated to overcome this problem. The total-order Sobol index is a measure of the contribution to the output variance of the $i$-th feature including all variance caused by the interactions with other features. It can be calculated as follows:

$$S_{Ti} = \frac{E_{\mathbf{X}\sim i}(V_{X_i}(Y \mid \mathbf{X}_{\sim i}))}{V(Y)} \tag{6}$$

Here $E_{\mathbf{X}\sim i}(V_{X_i}(Y \mid \mathbf{X}_{\sim i}))$ stands for the expected variance in model output $Y$ when all but the $i$-th feature are fixed.

### 2.4. Partial Dependence Plots

Once the importance of each input features has been determined with the Sobol indices, it is possible to visualise the type of relation between the input feature and the target feature. Partial Dependency Plots (PDP) [18] is an XAI method that can show whether the target variable and the input features have linear, monotonic, or more complex relationships. PDP's show the outcome of a model at a value of a feature when this value is substituted for all samples in the training set. PDP's divide the set of input features into to subsets: $C$ and $S$, where $S$ is the set of features that are investigated and $C$ is the set of the remaining features. PDP's marginalize the model output over the distribution of the features in set $C$, so that the plot shows the relation between the model output and the features in $S$. In practice the PD function is estimated by calculating the average output as displayed in the following Equation:

$$\hat{f}_s(x_s) = \frac{1}{n}\sum_{i=1}^{n} f(x_s, x_C^{(i)}) \tag{7}$$

Here $f$ is the trained machine learning model, $n$ the total number of points in the training set, $x_s$ the feature of interest, and $x_C^{(i)}$ are the feature values from the training set of the features that are not investigated. The output of this function expresses the average marginal effect of the feature $s$ on the prediction of the model.

### 2.5. Accumulated Local Effect plots

Accumulated Local Effect (ALE) plots [19] are a faster alternative to PDP's and in addition overcome the problem that PDP's assume feature independence. Just as with PDP's, ALE plots reduce the complex model function to a function that is dependent on only one or two features. The ALE plot, in contradiction to PDP's, show the change in prediction of a model over only a small window $z$ of a value of a feature. The difference in prediction can be seen as the effect of a feature on a single instance in a certain interval. The uncentered ALE function is defined as:

$$\hat{\tilde{f}}_{j,ALE}(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i:x_j^{(i)} \in N_j(k)} [\hat{f}(z_{k,j}, x_{\setminus j}^{(i)}) - \hat{f}(z_{k-1,j}, x_{\setminus j}^{(i)})] \tag{8}$$

Here, $\hat{f}$ is the machine learning model, $k$ the number of neighbourhoods, and $x_{\setminus j}$ all features except for feature of interest $j$. The sum on the right side of the equation adds the effect of all samples within interval or neighborhood $N_j(k)$. This sum is then divided by the number of samples in the interval in order to obtain the average difference of predictions in this interval. The left summation then accumulates all the averages over all intervals. The following equation then centers the effect such that the mean effect is $0$:

$$\hat{f}_{j,ALE,centered}(x) = \hat{\hat{f}}_{j,ALE}(x) - \frac{1}{n}\sum_{i=1}^{n}\hat{\hat{f}}_{j,ALE}(x_j^{(i)}) \tag{9}$$

The value of this function can be interpreted as the main effect of a feature at a certain value compared to the average prediction of the data. When the ALE function has, for instance, an estimate of $-4$ at $x_j = 2$ it indicates that when the $j^{th}$ feature has as value $2$ the prediction is lower by $4$ compared to the average prediction. The advantage of ALE plots is that they are not biased in a situation where features are correlated.

## 3. Data Acquisition and Processing

This section describes the data used and combined in this research to train the Automated Machine Learning Models with. The two sources used are; AIS data, explained in Section 3.1., and static ship data explained in Section 3.2.

### 3.1. AIS Data

AIS transceivers constantly transmit a vessel's identity, location, sailing speed, and course along with additional information about the destination and identity of the vessel. Most vessels are equipped with a Very High Frequency (VHF) AIS transceiver that allows local AIS data to be received and plotted on a chart plotter. Simultaneously, the transceiver sends out AIS data to other nearby receivers. The range of VHF receivers is approximately between 10-20 nautical miles. Due to the relatively short range of VHF receivers it is not possible to capture all worldwide AIS data with just one receiver. In order to overcome this issue, many third party applications have been developed that collect the AIS data from multiple local transponders and combine them in order to come up with a wider coverage. In collaboration with the Transferring Operational Data into Design Information for Ships project (TODDIS) [20], it was decided to use the free AISHub [21] as a third party AIS provider. AIShub aims to connect as many local AIS transceivers in order to obtain the best possible coverage. All connected stations are on land, or within coastal areas, meaning that there is no coverage on open ocean. The total coverage of AIShub can be seen in Figure 1.. The map shows a strong coverage in Europe, East coast North-America, and South-East Asia.

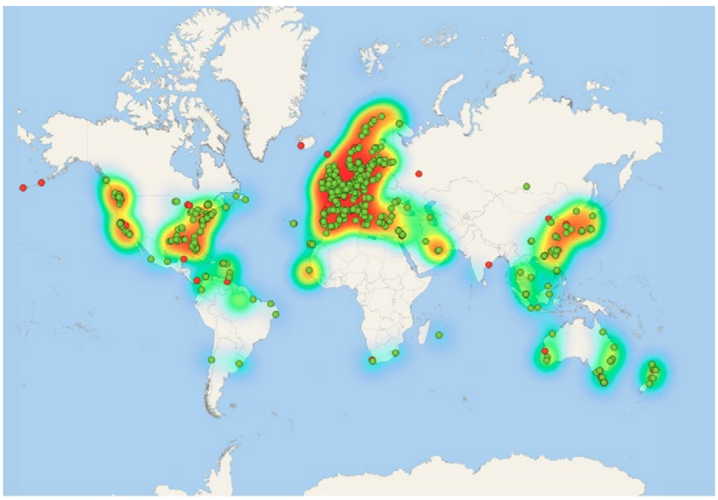


**Figure 1.**: AISHub coverage

The data collection process has started in September 2021 and is still going on at the moment of writing. At the moment of writing the Google Bigquery database contains approximately 9.3 billion rows with a total size of 1.5 TB. A schematic overview of the AIS data collection pipeline is given in Figure 2..

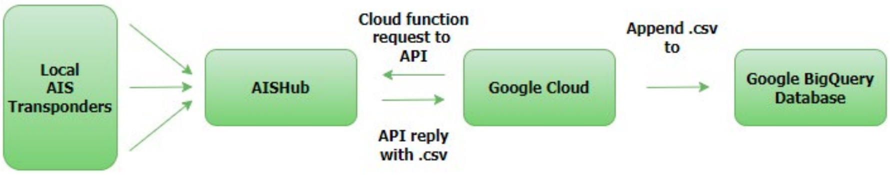



**Figure 2.**: Schematic overview of the AIS data collection pipeline

### 3.1.1. Data preprocessing

From this large set of raw AIS data, information can be extracted about a vessels operational performance capabilities. The columns of interest are: TSTAMP, Latitude, Longitude, COG, SOG, Heading, Navstat, IMO, A, B, C, D and Draught. The definition of these columns are given in Table 1 .. Furthermore, a subset of the data is selected so that only ocean going vessels are considered. This is done by only selecting vessels with a length of more than 80 meters, as recommended by a domain expert. Next, the samples with feature values that lie outside the allowed ranges shown in Table 1. are filtered out. Examples of this are samples that have a value of 511 for heading, which means that the heading is unknown. Any further samples that have unrealistic feature values, such as an SOG $> 40$ knots, are also removed. Finally, the navigational status of a vessel is checked for each record, and removed if the navigational status is 1 (at anchor) or 5 (moored). The resulting subset contains data from October up to February.

Even though the dataset spans a five month time window, many vessels have a relatively limited number of records in the dataset as shown in Figure 3.. In order to get a good grasp of the maximum operational performance capabilities of a vessel, a minimum number of data points per vessel is needed. Furthermore, these data points need to be from a variety of days. Simply looking at the data from one day might give a

Table 1.: Definition of features in raw AIS data

| Feature name | Explanation | Range |
|---|---|---|
| **MMSI** | Maritime Mobile Service Identity | N/A |
| **TSTAMP** | Timestamp in UTC date/time format | N/A |
| **Longitude** | Geographical longitude in degrees | [-180,180] |
| **Latitude** | Geographical latitude in degrees | [-90,90] |
| **COG** | Course over ground | [0,360] |
| **SOG** | Speed over ground | >0 |
| **Heading** | Current heading of vessel at time of last message | [0,360] |
| **Navstat** | Navigational status, indicates what operation the ship is performing | [0,15] |
| **IMO** | IMO ship identification number | N/A |
| **Name** | Ship name | N/A |
| **Type** | Vessel Type | [1,99] |
| **Callsign** | Vessel callsign | N/A |
| **A** | Distance from transceiver to bow in meters | N/A |
| **B** | Distance from transceiver to stern in meters | N/A |
| **C** | Distance from transceiver to port in meters | N/A |
| **D** | Distance from transceiver to starboard in meters | N/A |
| **DRAUGHT** | Draught of vessel in meters | >0 |
| **DEST** | Vessel destination | N/A |
| **ETA** | Estimated time of arrival in UTC date/time format | N/A |

biased view of a vessels performance as the conditions on that day might be especially dis- or advantageous. By trial and error it is determined that vessels in the dataset that have transmitted less than 300 data points, or on less than 4 unique days will be removed from the dataset. As only data points that have an AIS status different from 1, or 5 are selected from the Bigquery database, it is guaranteed that these data points are not from ships laying still in harbor.

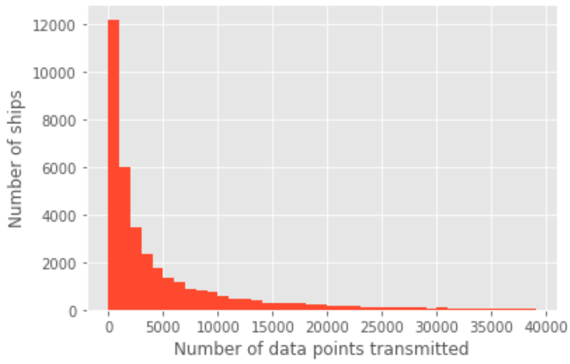



**Figure 3.**: For many vessels the number of logged data points is relatively low. Similarly, for many vessels the logged data points have been transmitted over a small number of days.

### 3.1.2. Feature extraction

The resulting dataset is 9 GB, has 197644394 rows and 12 columns describing the behaviour of 35645 unique vessels over a period of 5 months. The next step is to derive features from the AIS data that give information about the operational performance capabilities of a ship. This is done by calculating the following features for each vessel.

**Max speed:** The maximum SOG that a vessel has sailed. This feature is of interest to a naval architect because it shows how fast a ship can potentially sail. As SOG is part of the AIS data string, it can be simply extracted directly for each ship in the dataset.

**Median cruising speed:** This feature gives information about the median speed of a ship, and is thus less dependant on whether the collected data points come from ships that are sailing at full power. The median cruising speed is calculated by taking the median of all speed values that lie within the range

$$x \cdot SOG_{max} < SOG < SOG_{max} \quad \text{with} \quad x \in \{0.6, 0.7, 0.8, 0.9\} \tag{10}$$

This range is used in order to disregard data points that are recorded while ships are sailing in ports, or other waters where their movements are restricted.

**Max rotation:** The maximum rotation in degrees that a vessel has made in a time window of one minute. The feature is of interest for a naval architect because it is a quantification of the maneuverability of a ship. Rotation is not part of the raw AIS data, and thus needs to be derived. The rotation per minute can be calculated from AIS data by taking the difference in heading between two consecutive AIS records, dividing the difference in heading by the difference in time in seconds and then multiplying this by 60. An example edge case exists when the heading of the first record is 5, and the heading of the second record is 355. The difference between these two records is 350, while the ship has probably only rotated by 10 degrees. If the difference between the two records is larger than 180, the difference in heading is subtracted from 360 to account for the edge cases where the heading of two records are both close to 360 and 0. The formula used for calculating the rotation between two records is:

$$\Delta_t = TSTAMP_2 - TSTAMP_1 \tag{11}$$

$$\Delta_h = \begin{cases} 360 - |h_1 - h_2|, & \text{if } |h_1 - h_2| > 180 \\ |h_1 - h_2|, & \text{otherwise} \end{cases} \tag{12}$$

$$rotation = \frac{\Delta_h}{\Delta_t} \times 60 \tag{13}$$

Here $\Delta_t$ stands for the difference in time, $\Delta_h$ for the difference in heading, $h_1$ for the heading of record 1 and $h_2$ for the heading of record 2. Using consecutive AIS record that are no further than two minutes apart resulted in the best results.

Next to these target features, some additional input features are calculated from the AIS data. For each target feature, the draught of the vessel at the moment that the vessel produced the target value is logged. Furthermore, for each vessel the number of tugboats in the proximity is calculated at the moment the maximum rotation value is logged. This value is calculated by using the latitude and longitude coordinates in the AIS data to check within a radius of 600 meters, in a time interval of 1 minute before the max rotation is logged, and 1 minute after the max rotation is logged if there are vessels present that transmit AIS data with ship type 52, as this is the vessel code for tugboats.

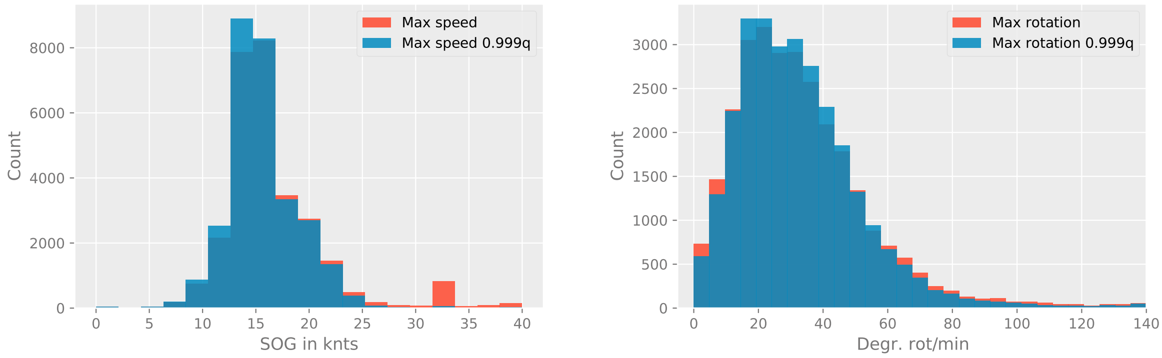



**Figure 4.**: Distributions of features calculated from AIS data. Taking the absolute maximum value seems more susceptible to outliers than taking the 0.999 quantile value.

The calculation of aforementioned target features can be influenced by missing values, or noise which is naturally present in AIS data [22]. In order to account for outliers that occur as a result of this, the 0.999 quantile value for each feature is also calculated. Figure 4. shows the distributions of the target features for both the maximum value, and the 0.999 quantile value. As expected, most of the 0.999 quantile values seem to have a slightly lower value than the absolute maximum values. This suggests that using the absolute maximum value is more susceptible to outliers. In Section 4. the feature that yields the best result is determined.

### 3.2. Static data

The second branch of data used in this research is static ship data. The dataset with static ship data comes from earlier work [5] where the data is provided by a reference database from C-Job Naval Architects. The dataset contains 230851 unique vessels, each described by 129 features. Examples of features are Length Overall (LOA), Depth (T), Breadth Overall (B), Dead Weight (DWT), Gross Tonnage (GT), Froude Number (Fn), but also more ship specific features like number of containers (TEU), or Bale cargo volume. Beause some features are more specific for certain ship types, not all features contain a value for each ship. After calculating the performance features from the AIS data as explained in Section 3.1., the instance is merged with additional information about the ship design from the reference database. Each instance in the resulting dataset now contains the features calculated from the AIS data, as well as the static ship information describing the ship's design. As each row now describes the performance characteristics and design of a unique vessel, the dataset has reduced much in size, containing only 27343 rows.

### 3.3. Imputing Missing values

The reference database with static ship data, discussed in Section 3.2., contains many columns with missing values. A missing value can occur in two cases; if the feature is not applicable to a ship, the feature "number_of_lorries"for instance, is only applicable for vehicle carrier ships and not for other ship types. In order to take account for this, for all ship types only the features that have a missing value percentage smaller than a predetermined threshold are selected. After manually inspecting the number of missing values for each ship type, the percentage is set to 30%. In almost all cases this percentage caused all features that are applicable to a ship to be selected, without selecting inapplicable features.

The second case where missing values can occur is when the value is simply unknown. If a feature value is unknown it can be derived from similar samples in the dataset. In order to impute the missing value the $k$-Nearest Neighbour (KNN) imputing method [23] can be applied. This method assigns each sample to $k$ neighbours based on the Euclidean distance between the feature values and then fills in the missing value as the mean feature value of the $k$ neighbours. For categorical features, the feature values have been transformed to numerical values.

## 4. Experiments & Results

In the experiments data from Container ships, General Cargo Ship Tween deck (GC ship tween deck), Multi Purpose General Cargo Ship (MP general cargo ship), and Roll-on Roll-off cargo ships (Ro-Ro cargo ships) is analysed. Before the XAI and GSA techniques can be applied, a Multi-Output AutoML model is trained which is capable of predicting the speed and rotating target features. After training, in three experiments, XAI and GSA techniques are used to investigate which features have the most influence on the outcomes of the the machine learning models.

### 4.1. Multi-output AutoML model

A multi-output regressor is trained that predicts for one ship type all target features. The AutoML pipeline is implemented in *Auto-sklearn 2.0* which requires the setting of only two hyperparameters; the time limit in seconds for finding appropriate models, and the per run time limit which is the time limit for a call to a single machine learning pipeline. The time limits are set as recommended in the documentation to respectively 7 hours, and 35 seconds. The dataset is split into a $70/30$ train test split and the $r^2$ scores on the test set are reported in Table 2.

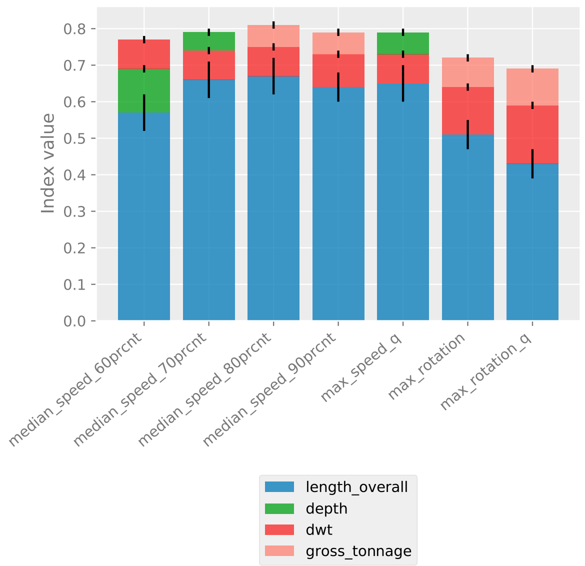
Table 2.: $r^2$ score of AutoML models on test set.

| | **median cruising speed** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **0.6** | **0.7** | **0.8** | **0.9** | **max speed 0.999q** | **max rot.** | **max rot 0.999q** |
| **Container ship** | 0.52 | 0.67 | 0.72 | 0.68 | 0.67 | 0.52 | 0.56 |
| **GC ship tween deck** | 0.59 | 0.60 | 0.56 | 0.54 | 0.56 | 0.32 | 0.31 |
| **MP GC ship** | 0.52 | 0.47 | 0.39 | 0.32 | 0.34 | 0.27 | 0.26 |
| **Ro-Ro cargo** | 0.82 | 0.81 | 0.83 | 0.81 | 0.82 | 0.46 | 0.51 |

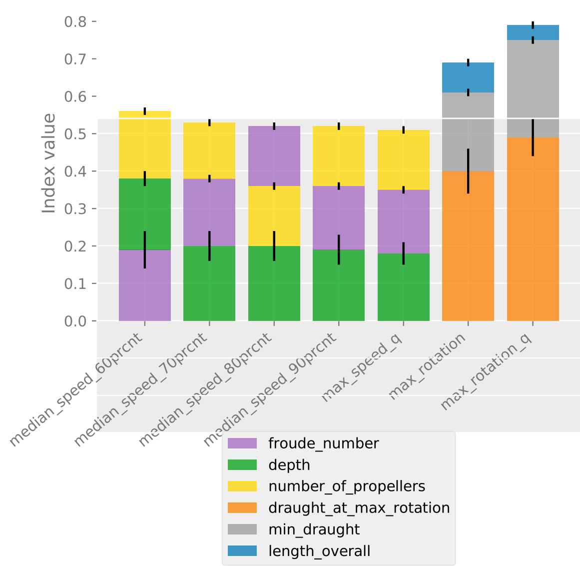
### 4.2. Sobol indices

The models described in Section 4.1. work in a black-box fashion, it accepts input and produces output, but it is not clear to the user how the model makes decisions in order to produce the output. In order to gain insight in which ship design parameters have influence on the performance of the entire fleet it is important to understand the working of the regression models. This subsection shows which input features explain the largest part of the variance of the model output, and thus which input features are most important for determining the value of the target features. For each ship type and target feature the first order, and total order Sobol index are calculated in order to gain insight in the effect of a feature with and without interaction with other features. As there exists very little difference between the resulting first order, and total order index, only the total order index is reported in Figure 5..

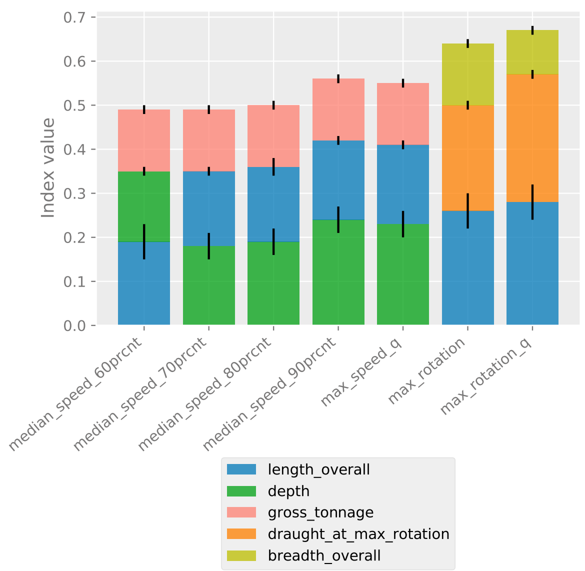
The plots show that for all target features for ship type Container ship the input feature *length overall* has the highest total order Sobol index. Although with lower index values than Container ships, GC ship tween decks has *depth* as the highest ranked Sobol index for most speed related target features. Furthermore, the draught related features obtain the highest Sobol index values for the target features related to rotation for this ship type. MP general cargo ships show a less clear picture of which features have the highest indices, but in most cases *length_overall* is the most important input feature for both speed and rotation related target features. For Ro-Ro cargo ships, the Sobol indices show that *length_overall* is the most important feature for all speed related target features, while also being one of the most important for both rotation related target features.



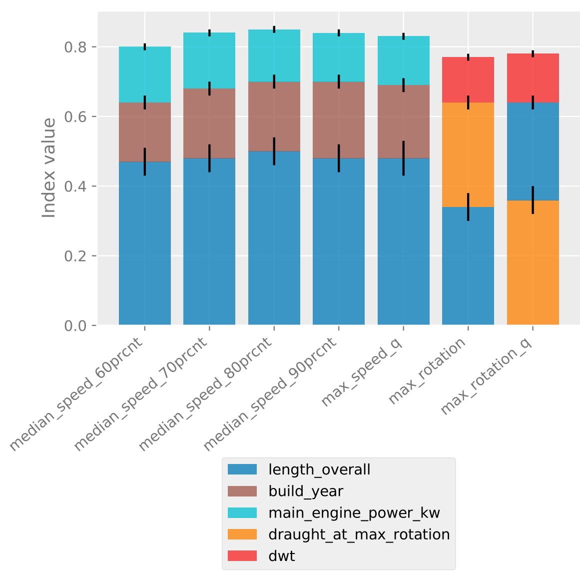

(a) Container ship




(b) GC ship tween deck




(c) MP general cargo ship




(d) Ro-Ro cargo ship

**Figure 5.**: Total order Sobol indices and confidence intervals for all ship types and target features. Sample size used for calculating the indices is $2^{12}$.

### 4.3. ALE plots

The Sobol indices give insights in which input features have the largest influence on the model output variance. In order to investigate in which way these features influence the model output, ALE plots are created for the feature with the highest ranked total order Sobol index. The plots show what kind of relation exists between changing the value of an input feature and the effect of the model prediction on the target feature. For all ship types and target features the influence of input parameters on the prediction of the AutoML model created in Section 4.1. is visualised by creating ALE plots.

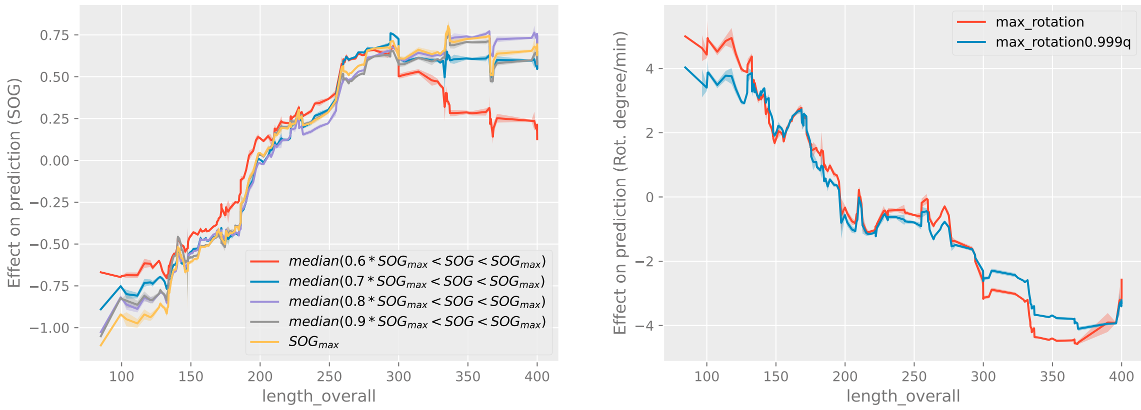



**Figure 6.**: ALE plots for Container ships for input feature *length overall*. Shaded areas are the $95\%$ confi-dence intervals calculated over the model predictions.

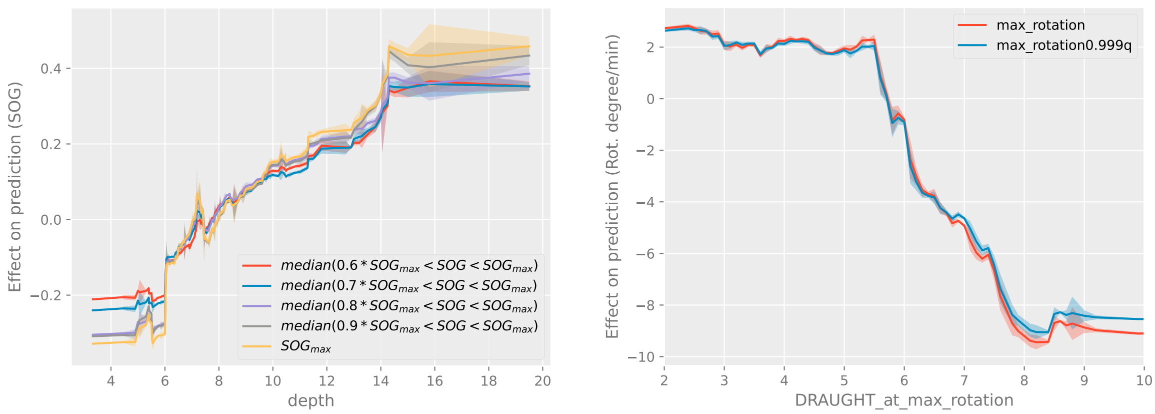



**Figure 7.**: ALE plots for GC ship tween decksfor *depth* (left) and *draught at max rotation* (right). Shaded areas are the $95\%$ confidence intervals calculated over the model predictions.

Figure 6. shows that there exists a positive trend between the overall length of container ships and the speed related target features. The plots show an oscillating line with a confidence interval, as ALE plots calculate an average of the target feature over samples that lie in a small interval. What can also be noticed is that there exists little difference between the intercept of the ALE lines. This can be attributed to the fact that the y-axes for ALE plots show only the effect on prediction. The ALE plot for rotation related features shows that the predicted value of degrees of rotation per minute decreases when the overall length of a ship increases.

In Figure 7. the plots show that for GC ship tween decks there exists a positive trend between the predicted speed related target features and the depth of the ship, although the effect is very small as it ranges from $-0.3$ to $0.4$ over the complete range of depth values. The ALE plot for rotation related target features shows little to no effect on the prediction for ships with a draught at max rotation between $2$ and $6$

meters. Between 6 and 8 meters there is a sharp drop in effect on predicted degrees of rotation per minute. From 8 to 10 meters there is again little to no effect on the predicted target value.

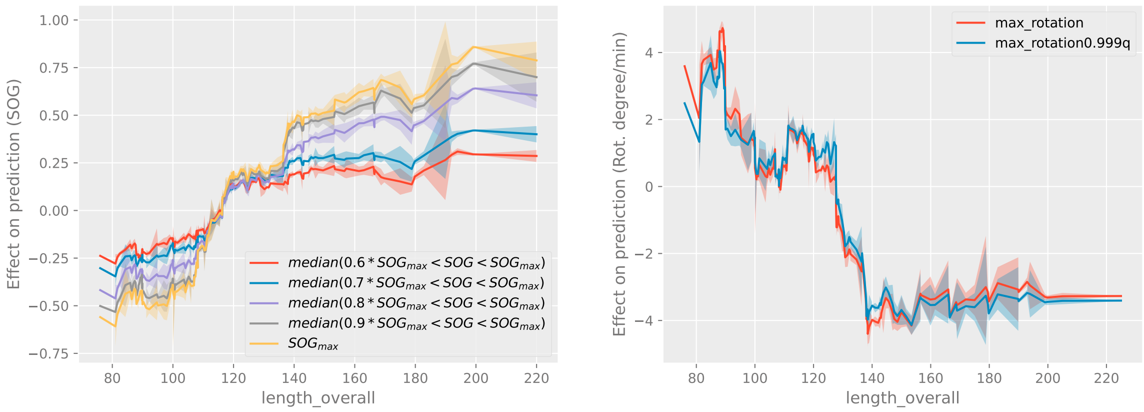



**Figure 8.**: ALE plots for MP general cargo ships with input feature *length_overall*. Shaded areas are the 95% confidence intervals calculated over the model predictions.

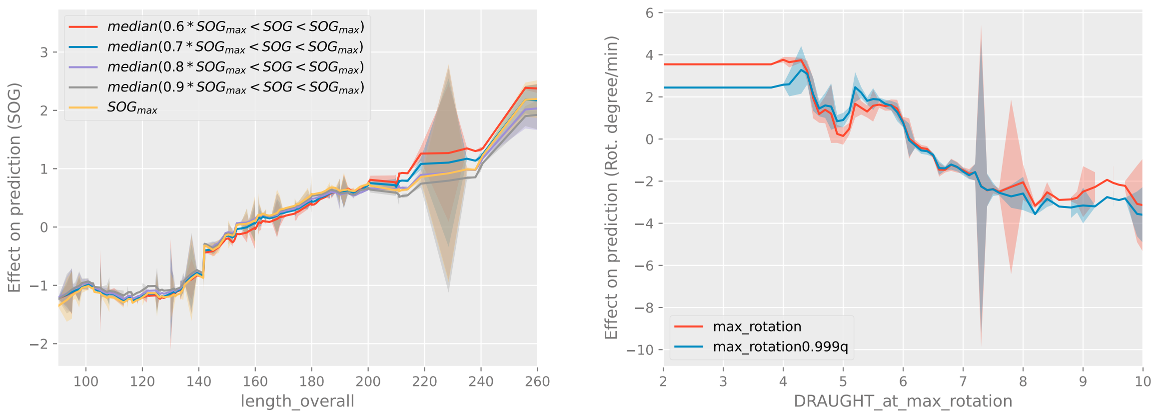



**Figure 9.**: ALE plots for Ro-ro cargo ships for input feature *length overall* (left) and *draught at max rotation* (right). Shaded areas are the 95% confidence intervals calculated over the model predictions.

The ALE plots for Multi Purpose general cargo ships can be found in Figure 8.. Here the plots again show a positive trend between the overall length of a ship and the predicted target value for speed related target features. Furthermore, it can be noticed that for rotation related features the overall length of a ship has a negative effect on predicted degrees of rotation per minute. Left subfigure i n F igure 8 . shows a relatively large difference between ships with an overall length between 80 and 140 meters. For ships longer than 140 meters the plot shows that there is little to no difference in prediction.

Figure 9. shows the ALE plots for Ro-ro Cargo ships, where the left subfigure s hows a p ositive trend between the overall length of the ships and the speed related target features, while the right subfigure shows a negative trend between the predicted degrees of rotation per minute and the draught at max rotation of a ship.

### 4.4. Partial dependence plots

The influence of input features on the model output can also be visualised with PDP's. In contrast to ALE plots, PDP's show the absolute predicted value on the y-axis instead of the effect on prediction. Figure 10. shows the PDP's for container ships for both speed related and rotation related target features.

The features plotted on the x-axis are the highest ranked total order Sobol indices as shown in Figure 5.. In the left subfigure of Figure 10. the PDP shows that there is a positive correlation between the overall length of a vessel and the speed related target features, just as the ALE plots showed. Right subfigure of Figure 10. shows a negative correlation between the overall length and rotational capabilities of a vessel, which is also similar to what the ALE plots showed. In order to compare the influence of the overall length on the maximum speed of all vessels simulatenously, the PDP plots are merged together in Figure 11. The plot shows that GC ship tween deck, and MP cargo ship are shorter vessels and that their length has a lesser impact on the $SOG_{max}$ than Container ships and Ro-Ro cargo ships.

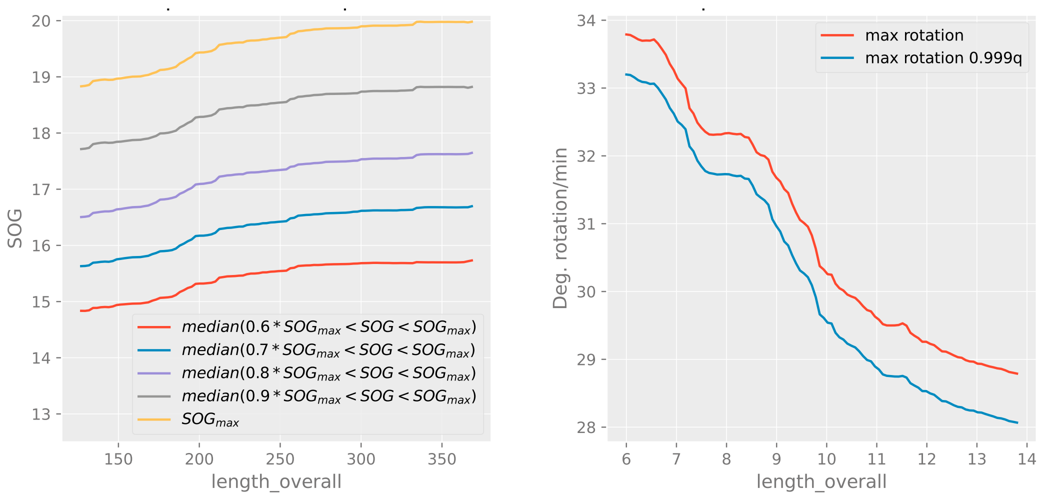



**Figure 10.**: PDP's for container ships for input feature *length_overall* for the speed related target features (left) and the rotation related target features (right).

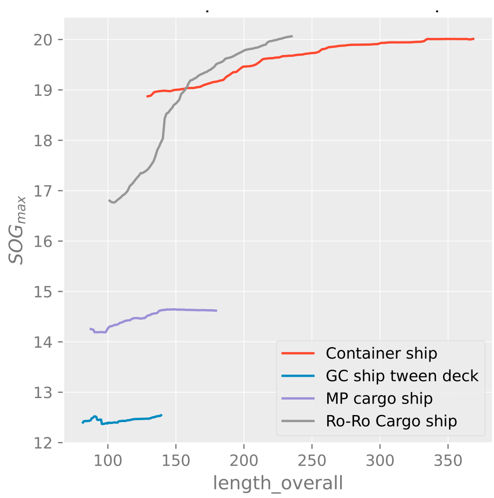



**Figure 11.**: PDP for all investigated ship types showing the influence of the overall length of a vessel on the prediction of $SOG_{max}$.

## 5. Discussion & Conclusion

The experiments in Section 4.1. showed how for Container ships, MP cargo ships, GC ship tween deck, and Ro-ro cargo ships regression models can be created for speed and rotation related target features. In some cases taking the absolute maximum value for a target feature was susceptible to outliers. In order to take account for this, the 0.999th quantile value is used. Analyzing the multi-output AutoML model with GSA and XAI techniques shows how the multiple targets change to a permutation of the input features. The Sobol indices for these models showed that for Container ships, the overall length of the ship is the most important input feature for both speed related and rotation related target features. For GC ship tween deck and Ro-ro cargo the results also show that the overall length is the most important input feature for the speed related target features. For MP general cargo ships the depth is the most important input feature for the speed related target feature, while the overall length is the most important feature for rotation. For GC ship tween deck, and Ro-ro cargo ships the Sobol indices showed that the draught at max rotation is the most important input feature.

The GSA and XAI methods used in Section 4., show how the input features influence the output of the model, and thus how the design parameters influence the speed and rotational capabilities of a ship. The ALE plots and PDP's showed a positive trend between the predicted speed capabilities and overall length of ships. An increase of the overall length of a ship often means that the breadth of the ship increases as well, but only up to a certain level, as ships often have to be narrow enough to fit through locks used in busy sailing routes. Increasing the width of a ship is associated with the largest increase in resistance called wave resistance. On the other hand, increasing the length of a ship is associated with a much smaller increase in resistance called boundary layer induced friction [24]. To account for the increase in resistance for longer and wider ships, these vessels are often installed with extra engine power. Even though the engine power was one of the input features, it is never selected as most important input feature. As the breadth of Container ships tend to max out around 60 meters, all extra engine power installed after this can be used to compensate for extra length of a ship. As the extra resistance as a result of a longer ship is much smaller than that of a wider ship, the results show that longer container ships exploit the extra engine power in a more efficient way than shorter or broader ships. The influence of the overall length of a vessel on the rotational capabilities of Container ships can be explained by the fact that a shorter ship has to displace less water when turning than a longer ship. These explanations of the importance of the overall length of a ship are confirmed by the ALE plots and PDP's which show that there exists a negative trend between the rotational capabilities of Container ships and the overall length.

A limiting factor in this research was that some important information, such as accurate weather data, or information about a ships engine settings was not available. Furthermore, the coverage of AIS hub is mainly focused on coastal waters, and does not cover the open oceans between the continents. In future work, using a source of AIS data that has more coverage might increase the quality of the data and thus the performance of the regression models.

## 6. References

[1] Ziaul Haque Munim, Mariia Dushenko, Veronica Jaramillo Jimenez, Mohammad Hassan Shakil, and Marius Imset. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Maritime Policy & Management, 47(5):577–597, 2020.

[2] Zhong Ming, Min De quan, and Liu Jian qiu. A dynamic vessel fleet planning model in uncertain circumstances. In 2009 IEEE International Conference on Automation and Logistics, pages 379–384, 2009.

[3] JP Panda. Machine learning for naval architecture, ocean and marine engineering. arXiv preprint arXiv:2109.05574, 2021.

[4] Henrique M Gaspar, Donna H Rhodes, Adam M Ross, and Stein Ove Erikstad. Addressing complexity aspects in conceptual ship design: A systems engineering approach. Journal of Ship Production and Design, 28(04):145–159, 2012.

[5] R de Winter, B van Stein, THW B¨ack, and V Bertram. Ship design performance and cost optimization with machine learning. COMPIT'21, pages 185–196, 2021.

[6] Robert G Keane, Laury Deschamps, and Steve Maguire. Reducing detail design and construction work content by cost-effective decisions in early stage naval ship design. In SNAME Maritime Convention. OnePetro, 2014.

[7] Yu Ao, Yunbo Li, Jiaye Gong, and Shaofan Li. An artificial intelligence-aided design (aiad) of ship hull structures. Journal of Ocean Engineering and Science, 2021.

[8] Tomasz Abramowski. Application of artificial intelligence methods to preliminary design of ships and ship performance optimization. Naval Engineers Journal, 125:101–112, 09 2013.

[9] Vaibhav Kumar and ML Garg. Predictive analytics: a review of trends and techniques. International Journal of Computer Applications, 182(1):31–37, 2018.

[10] Fernandez Arguedas V, Pallotta G, and Vespe M. Maritime traffic networks. from historical positioning data to unsupervised maritime traffic monitoring. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 19(3):722–732, 2018.

[11] Sepideh Jafarzadeh and Ingrid Schjølberg. Operational profiles of ships in norwegian waters: An activity-based approach to assess the benefits of hybrid and electric propulsion. Transportation Research Part D: Transport and Environment, 65:500–523, 2018.

[12] Morten Winther, Jesper H. Christensen, Marlene S. Plejdrup, Erik S. Ravn, ´Omar F. Eriksson, and Hans Otto Kris-tensen. Emission inventories for ships in the arctic based on satellite sampled AIS data. Atmospheric Environment, 91:1–14, July 2014.

[13] Kenneth Mitchell and Brandan Scully. Waterway performance monitoring with automatic identification system data. Transportation Research Record Journal of the Transportation Research Board, 2426:20, 12 2014.

[14] Longbiao Chen, Daqing Zhang, Xiaojuan Ma, Leye Wang, Shijian Li, Z. Wu, and Gang Pan. Container port performance measurement and comparison leveraging ship gps traces and maritime open data. IEEE Transactions on Intelligent Transportation Systems, pages 1–16, 12 2015.

[15] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. Automatic Machine Learning: Methods, Systems, Chal-lenges. Springer, 2019.

[16] Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, and Frank Hutter. Auto-sklearn 2.0: The next generation. CoRR, abs/2007.04074, 2020.

[17] Jiri Nossent, Pieter Elsen, and Willy Bauwens. Sobol'sensitivity analysis of a complex environmental model. Environmental Modelling & Software, 26(12):1515–1525, 2011.

[18] Christoph Molnar, Timo Freiesleben, Gunnar K¨onig, Giuseppe Casalicchio, Marvin N. Wright, and Bernd Bischl. Relating the partial dependence plot and permutation feature importance to the data generating process, 2021.

[19] Daniel W. Apley and Jingyu Zhu. Visualizing the effects of predictor variables in black box supervised learning models, 2016.

[20] NHL Stenden. Toddis project. https://www.nhlstenden.com/nieuws-en-artikelen/toddis-project-gebruik-van-big-data-voor-innovatieve-scheepsontwerpen. Ac-cessed: 17-05-2022.

[21] AISHub. AISHub webpage. https://www.aishub.net. Accessed: 25-03-2022.

[22] Kwang Il Kim and Keon Myung Lee. Mining of missing ship trajectory pattern in automatic identification system. International Journal of Engineering & Technology, 7(2.12):167–170, 2018.

[23] Padraig Cunningham and Sarah Jane Delany. k-nearest neighbour classifiers: 2nd edition (with python examples). CoRR, abs/2004.04523, 2020.

[24] Qingsong Zeng, Robert Hekkenberg, Cornel Thill, and Hans Hopman. Scale effects on the wave-making resis-tance of ships sailing in shallow water. Ocean Engineering, 212:107654, 2020.