

Article

Pairwise-Comparison Based Semi-SPO Method for Ship Inspection Planning in Maritime Transportation

Ying Yang ¹, Ran Yan ^{2,*} and Hans Wang ²¹ Logistics and Transportation Division, Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China² Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong 999077, China

* Correspondence: ran-angel.yan@polyu.edu.hk

Abstract: Port state control (PSC) plays an important role in enhancing maritime safety and protecting the marine environment. Since the inspection resources are limited and the inspection process is costly and time-consuming, a critical issue for port states to guarantee inspection efficiency is to accurately select ships with a high risk for inspection. To address this issue, this study proposes three prediction models to predict the ship deficiency number and a ship selection optimization model based on the prediction results to target the riskiest ships for inspection. In addition to a linear regression model for ship deficiency number prediction solved by the least squares method, we establish two prediction models with the pairwise-comparison target based semi-“smart predict then optimize” (semi-SPO) method. Specifically, a linear programming model and a support vector machine (SVM) model are built and both have a loss function to minimize the sum of predicted ranking errors of each pair of ships regarding their deficiency numbers. We use the Hong Kong port as a case study, which shows that the SVM model based on the semi-SPO approach performs best among the three models with the least computation time and best ship selection decisions.



Citation: Yang, Y.; Yan, R.; Wang, H. Pairwise-Comparison Based Semi-SPO Method for Ship Inspection Planning in Maritime Transportation. *J. Mar. Sci. Eng.* **2022**, *10*, 1696. <https://doi.org/10.3390/jmse10111696>

Academic Editor: Claudio Ferrari

Received: 7 October 2022

Accepted: 24 October 2022

Published: 8 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: port state control; high-risk ship selection; smart predict then optimize; pairwise-comparison loss function

1. Introduction

Maritime transportation is the backbone of international trade and the global supply chain [1]. However, the loss of maritime accidents can be huge because it always brings great casualties and financial losses [2]. In recent years, sustainable shipping has also attracted worldwide attention since maritime transportation produces heavy greenhouse gas emissions and pollutants, which contribute significantly to global climate change [3]. Therefore, growing awareness has been given to the safety and sustainability of maritime transportation. To effectively improve ship safety and reduce pollutants in shipping and port activities, numerous international conventions and regulations have been implemented by the International Maritime Organization (IMO) [4,5] and local governments [6]. Generally, a ship is regarded as substandard if its conditions, crew, or operations are substantially below the standards required by these regulations [7]. There are mainly two guidelines to defend against substandard shipping. The first one is the flag state of a ship, which is the nationality or tribe a ship belongs to, and under whose laws the ship is plying in international waters. However, it is believed that the flag states cannot effectively identify the substandard ships because the ships may not visit the ports under their flag state regularly [8]. In this context, PSC inspection is implemented over the world [9], which is developed for ports to inspect foreign visiting ships and to ensure that the ships' conditions are in accordance with the international conventions. If any part of a ship does not meet the specific requirements of the conventions, the relevant ship deficiency will be recorded. Moreover, if too many or critical ship deficiencies are detected on the ship, the port state

may take an intervention action preventing the ship from proceeding to the sea, i.e., detention. To achieve efficient inspection, Memorandums of Understandings (MoUs) on PSC are established and signed by neighboring countries and regions, within which the policies and standards of ship selection and inspection are uniform.

Before carrying out one day's PSC inspections, one critical issue faced by the port authority is to select the high-risk ships, i.e., the ships that have the most deficiencies or have the highest risk of detention, for inspection from all the foreign visiting ships because the inspection resources are limited and inspecting all visiting ships are both costly and time-consuming. Therefore, many ship selection schemes have been adopted around the world. For example, in the Tokyo MoU, New Inspection Regime (NIR) is applied to determine ship inspection priority by calculating the ship risk profile (SRP) [10]. These ship selection regulations currently used in practice are easy to implement but have obvious disadvantages that may reduce the efficiency of ship inspection [11]. Therefore, a considerable number of studies focus on developing more advanced ship selection models to select ships with a larger number of deficiencies or with higher detention probability from visiting ships. Most of these studies follow a predict-then-optimize framework by first predicting the risk of a ship and then selecting ships for inspection according to the prediction results, i.e., ship deficiency number or detention probability, to maximize the inspection benefits or minimize the total inspection cost [8,12]. However, while these studies have made great efforts in improving the accuracy of the prediction model, they may fall into the trap that a more accurate prediction model does not necessarily guarantee better ship selection decisions. This is because that their model does not account for how the predictions will be used in the optimization model, and thus may lead to sub-optimal decisions [13]. In the ship selection problem, the performance of the ship selection optimization model is highly related to the relative ranking of ships regarding their ship deficiency number rather than their absolute deficiency numbers. There are methods that fully integrate prediction and optimization, called "smart predict then optimize" (SPO) [13], which, however, are usually computational intractable. In fact, an alternative approach is to partially integrate prediction and optimization models in a way such that not too much computation time and resources are cost [14], and this method is called semi-"smart predict then optimize" (semi-SPO) approach. Although SPO can directly leverage the optimization problem structure, i.e., objective and all constraints, for designing better prediction models, this method is not quite applicable to our problem, since it requires predicting the ranking of all ships coming to the port based on ship deficiency number and thus costs too much time. By contrast, the semi-SPO approach can be more time-efficient by considering part of the information of the ship selection optimization model into the prediction model. Therefore, our study aims to improve the effectiveness of the ship selection scheme by establishing a semi-SPO approach, where the prediction model of ship deficiency number partially takes the structure of the subsequent ship selection optimization model into account to improve decision quality.

Specifically, we try to take the structure of the ship selection optimization model into account and minimize the prediction error regarding the relative ranking of each ship pair considering their numbers of deficiencies. We evaluate the ship selection decisions based on the proposed semi-SPO prediction models and compare them with the classic model which aims to minimize the total prediction error using the real inspection data at the port of Hong Kong. The remainder of this paper is organized as follows: Section 2 reviews the literature on the ship selection scheme for inspection. Section 3 establishes a mathematical model for the ship selection problem and describes the data used in this study. Section 4 introduces the prediction and optimization models. Section 5 validates and compares the models using the data at Hong Kong port. Section 6 draws a conclusion and provides an outlook of the future research directions.

2. Literature Review

A recent literature review divides the research topics related to PSC inspection into four categories: factors influencing PSC inspection results, high-risk ship selection schemes, effects of PSC inspection, and suggestions to improve PSC inspection [15]. Our study focuses on the ship selection scheme for inspection, so we will mainly review the literature in this area.

A number of early research studies have been focused on improving the efficiency of the ship selection scheme for inspection. Li [16] attempts to establish a new system to classify risky ships based on their quality factors, e.g., age, flag and classifications. Similarly, Degré [17] proposes “Risk Concept” to help identify high risk vessels and to inspect them accordingly. There are also some studies that adopt machine learning models to address the ship selection problem. Xu et al. [18] present a risk assessment system based on SVM, which estimates the risk of each candidate ship using its generic factors and history inspection factors so as to select high-risk ones before conducting PSC inspection. Xu et al. [19] then improve their risk assessment system by including more target factors gained by the web mining technique. Based on these studies, Gao et al. [20] develop a new risk assessment system for PSC inspection, which uses a novel SVM and k-nearest neighbor (KNN-SVM) to remove noisy training examples and bag of words to extract some new target factors for the PSC inspection database. Chi and Jun [21] propose a new mathematical model with the advantages of automatic optimization and self-evolution, by using generalized additive modeling. However, all of the above models only consider static factors such as ship age, flag, and size when analyzing the risk score of a ship, while the dynamic factors of the ships, such as the change of flag, are not taken into account.

It was not until 2011 that NIR was forced in Paris MoU as the new ship selection scheme at the time. The NIR and the SRP were also implemented by the Tokyo MoU in 2014, which improves the efficiency of PSC inspection and brings positive effects to maritime safety, security, pollution prevention, and working conditions [22]. Nevertheless, the NIR also has some drawbacks [11]: First, the NIR only considers static ship-related features and limited historical ship inspection records when calculating ship risk, while other factors such as previous ship accidents are neglected. Second, the NIR adopts a simple weighted sum method to calculate the risk score where the weights of each feature are based on expert judgement without considering the real inspection practice, which may lead to biased results. In addition, the SRP roughly divides the ships into three risk categories which may be inaccurate to indicate the ship risk level. Therefore, although the influence of the NIR is generally positive, these drawbacks may hinder its effectiveness.

In recent years, more advanced and accurate ship selection models have been introduced to PSC inspection. Yang et al. [12] establish a data-driven Bayesian Network (BN) approach to analyzing risk factors influencing PSC inspections and predicting the probability of vessel detention. Specifically, they collect the inspection data of bulk carriers in seven major European countries from 2005 to 2008 in Paris MoU to obtain the relevant risk factors, which include the number of deficiencies, type of inspection, recognized organization (RO) and vessel age. Based on these risk factors, they predict the detention probabilities under different situations by BN. Yang et al. [23] also present a BN model to determine vessel detention rates after adding company performance as a new indicator in PSC inspection. Wang et al. [8] develop a data-driven BN classifier called the tree augmented naive Bayes classifier to identify high-risk foreign vessels coming to a port authority. They use the PSC inspection records at the Hong Kong port in 2017 as the data set and validate the effectiveness of their model compared to the SRP currently used in practice. Dinis et al. [24] establish an approach for risk assessment of individual ships and maritime traffic based on BNs. They construct the BNs based on static risk factors adopted by the Paris MoU NIR on PSC. In addition to BN models, Yan et al. [25] develop a classification model called balanced random forest to predict ship detention probability. By using the inspection records at the Hong Kong port for three years, they validate that their model outperforms the current SRP and is effective in predicting ship detention, which is not a trivial task as the low

detention rate leads to a highly imbalanced inspection records data set. There are also some studies that focus on combining past incidents and detention information for targeting high-risk vessels for inspection. For example, Heij and Knapp [26] present five combined methods to classify vessels based on these two risk dimensions, each of which involves extensive sets of factors. Knapp and Heij [27] further improve the proposed approach proposed by Heij and Knapp [26] and evaluate seven targeting methods against random selection of vessels using empirical data for 2018. They conduct numerical experiments on three comprehensive data sets that cover the world fleet and show a 14–27% improvement compared to random selection.

The above review shows that, in the most recent research, more comprehensive factors are considered to predict the ship deficiency number or probability of ship detention so as to support the ship selection for PSC inspection. Nevertheless, none of the previous models considers how these prediction results will be used in the ship selection process. Our study tries to bridge the research gap in the ship selection scheme and implements both the classic predict-then-optimize model and the semi-SPO model to address the significance of taking the structure of the subsequent ship selection optimization model into the prediction process.

3. Problem and Data Description

3.1. Ship Selection Problem

We first address the ship selection problem of PSC inspection in practice. According to the working process of the PSC authorities, in the morning of each day, a set of ships coming to the port state are candidates to be inspected by the PSCOs. However, inspecting all visiting ships is neither practical, due to limited resources, nor unnecessary, since some ships are in satisfactory condition. In fact, only ships with large deficiency number and/or high detention risk are worth inspection. Therefore, how to select the ships with the highest risk levels for inspection is a critical issue for PSC. In this study, we adopt ship deficiency number as the risk indicator because the detention rate is extremely low. Specifically, the ship selection problem aims to select S ships for inspection from the M foreign ships coming to the port on that day such that the total number of deficiencies identified on all inspected ships can be maximized. We note that the number of selected ships S is much smaller than the total number of visiting ships M .

The detailed notation used in our problem is provided in Table 1.

Table 1. Notation in the ship selection problem.

Notation	Definition
M	The number of foreign ships coming to the port on that day, indexed by m .
S	The number of selected ships for inspection, indexed by s .
n_m	The number of deficiencies of ship m .
u_m	A binary variable that equals 1 if ship m is selected for inspection, or 0, otherwise.

Mathematically, the ship selection problem can be modeled as follows:

$$\max_u \sum_{m=1}^M n_m u_m \quad (1)$$

$$\text{s.t.} \quad \sum_{m=1}^M u_m \leq S, \quad (2)$$

$$u_m \in \{0, 1\} \quad \forall m \in M. \quad (3)$$

Objective (1) maximizes the total number of deficiencies that can be detected from the ships selected for inspection. Constraint (2) guarantees that no more than S ships can be selected for inspection. Constraints (3) define the domains of variables. We note that

the deficiency number of each ship n_m is an unknown parameter before an inspection, and thus this model cannot be solved directly. In previous studies [23,25], they adopt a two-stage framework, where the first stage is to predict the ship deficiency number and the second stage is to solve the optimization model with the predicted deficiency number as input. In particular, the optimization is an integer linear model and can be directly fed into an off-the-shelf solver. However, their two-stage framework does not consider how the prediction results will be used in the subsequent optimization model. Therefore, we will try to solve this problem by establishing semi-SPO approaches, which are presented in Section 4.

3.2. Data Description

In this study, we collect ship inspection records from the Asia Pacific Computerized Information System (APCIS) (<https://apcis.tmo.u.org/> (accessed on 1 December 2021)) supported by the MoU and World Register of Ships (WRS) database (http://www.tokyo-mou.org/inspections_detentions/psc_database.php (accessed on 1 December 2021)). APCIS provides detailed inspection records at the Hong Kong port and PSC-related information of the inspected ships within the Tokyo MoU, while WRS provides ship-related factors such as ship dimension, registration, and ownership. We combine the two databases with a ship IMO number and form a data set which contains a total of 4404 inspection records together with ship specifications at the Hong Kong port from 2015 to 2019. One record contains ship inspection results (i.e., deficiency conditions and ship detention) and ship specifications (e.g., ship characteristics and performance of ship management parties). In the ship selection problem, we consider ship deficiency number as the ship risk indicator, which is denoted by n . Then, we select 14 features (where the feature vector is denoted by x) from the original data set which are highly related to the number of deficiencies a ship has. These features can be divided into three categories, namely ship-related features (including ship age, gross tonnage, length, depth, beam, and type), PSC-related features (including flag performance, recognized organization (RO) performance, and company performance over the past three years reported by the Tokyo MoU), and ship historical inspection features (including period since the last inspection, deficiency number of the last inspection, total detentions in all previous inspections, flag change times, and casualties in the last five years). Then, we randomize the whole data set and divide it into the training set, validation set, and test set with each containing 60%, 20%, and 20% of all records, i.e., 2643, 881, and 881 records, respectively.

The explanation and descriptive statistics of the features are presented in Table 2. From the best to the worst, the states of ship flag performance are ‘white’, ‘grey’, and ‘black’, and the states of ship RO and company performances are ‘high’, ‘medium’, ‘low’, and ‘very low’, respectively. Since some inspection records have empty features, we fill the missing values by the statistical values, which are gained from records in the training set with the valid corresponding feature. Specifically, for undefined flag performance, RO performance, or company performance, we fill them with the mode value of the corresponding feature; for empty ship length or depth value, we fill them with the mean value of the corresponding feature; for records with no last inspection information, we fill the period since the last inspection and deficiency number of the last inspection with the median or mode value of the corresponding features, respectively. Moreover, we encode the categorical features so that they can be used by machine learning models. To be more specific, for feature “type”, we adopt one-hot encoding and the ship type is 1 if this ship belongs to the specific type and 0, otherwise; for feature “flag performance”, we adopt label encoding and ‘white’, ‘grey’, ‘black’ are encoded into 3, 2, 1, respectively; for features “RO performance” and “company performance”, label encoding is also adopted, and ‘high’, ‘medium’, ‘low’, and ‘very low’ are encoded into 4, 3, 2, 1, respectively; for feature “casualties in the last five years”, it is encoded into 1 if any casualty occurs in the last five year and 0, otherwise.

Table 2. Feature explanation and descriptive statistics.

Feature Name	Meaning	Mean Value *	Max Value *	Min Value *
ship age (year)	The time interval in year between a ship's keel laid date and the current inspection date.	11.42	48.00	0
gross tonnage (100 cubic feet)	The internal volume of a ship.	43,388.43	217,612.00	299.00
length (meter)	The length of a ship.	212.59	400.00	40.75
depth (meter)	The vertical length of a ship from the top of the keel to the underside of the upper deck at side.	17.52	38.00	3.30
beam (meter)	The width of a ship.	31.49	63.10	7.80
type	Ship type searched from the database and according to the annual report on PSC from [28].	/	/	/
flag performance	The performance of ship flag state calculated by the historical performance of the ships under the flag over the past three years reported in [29].	/	/	/
RO performance	The performance of ship recognized organization determined by the inspection and detention history of its ships over the last three years [30].	/	/	/
company performance	The performance of ship company based on the Tokyo MoU database considering its ships' inspection performance in a running 36-month period [10].	/	/	/
period since the last inspection (month)	The time interval in month between the last initial inspection date and the current inspection date of a ship.	8.56	178	0
deficiency number of the last inspection	The number of deficiencies of a ship identified in the last initial PSC inspection within the Tokyo MoU.	3.28	55	0
total detentions in all previous inspections	The total detention times of a ship in all PSC inspections by all PSC authorities.	0.70	19	0
flag change times	The total flag changing times of a ship since its keel laid date.	0.73	8	0
casualties in the last five years	Whether a ship has encountered casualties in the last five years.	/	/	/

*: These three columns represent the average, maximal, and minima values of numerical features of ships in the entire data set, respectively.

4. Prediction and Optimization Approaches

To solve the ship selection problem described in Section 3.1, we first propose three different prediction models to predict the number of deficiencies of each ship. Then, we develop an optimization model to select ships with the highest risk for inspection based on the predicted deficiency numbers.

4.1. Prediction Model

4.1.1. M1: Linear Regression Model with Classic Loss Function

One intuitive method to predict the ship deficiency number is linear regression calculated by a least squares method, which is a standard approach in regression analysis to approximating the target value n by minimizing the sum of squared errors between the predicted value \hat{n} and the actual value n . In model M1, the relationship between the input values x and the target value n is modeled using a linear regression model, where the best fitting coefficients are obtained by least squares functions.

Specifically, given a training data set with N inspection records denoted by $(x_1, n_1), (x_2, n_2), \dots, (x_N, n_N)$, where $x_i, i \in \{1, \dots, N\}$ is the input feature vector of a ship, and n_i is the target value (i.e., deficiency number) of data record i . We assume that n_i can be predicted using a linear regression model with input feature vector x_i as shown below:

$$\hat{n}_i = w^T x_i + b, \quad (4)$$

where w is the weighted vector and b is a constant term. These coefficients can be obtained by the least squares method presented below:

$$\begin{aligned} (w_{LR}^*, b_{LR}^*) &= \arg \min_{(w,b)} \sum_{i=1}^N (\hat{n}_i - n_i)^2 \\ &= \arg \min_{(w,b)} \sum_{i=1}^N (w^T x_i + b - n_i)^2. \end{aligned} \quad (5)$$

Then, given a new test sample with feature vector x_0 , its predicted deficiency number \hat{n}_0 is $(w_{LR}^*)^T x_0 + b_{LR}^*$.

4.1.2. M2: Linear Programming Model with a Pairwise-Comparison Target

One obvious drawback of M1 is that the linear regression model for ship deficiency number prediction does not consider the subsequent optimization model for ship selection, whose core concern is the relative ranking of ships considering their deficiency numbers, rather than the absolute deficiency numbers of the ships predicted by the linear regression model. Consequently, a better prediction model of ship deficiency number in terms of prediction error might not necessarily induce a better decision.

We take the example in Table 3 to demonstrate this pitfall. In this example, three ships coming to the port are candidates for inspection and three linear regression models (namely LR1, LR2, and LR3) are developed to predict ship deficiency number. The first three rows of column #Actual show the actual deficiency number of each ship, while the last two rows of column #Actual present the best ship selection decision under the actual deficiency number. The row MSE shows MSE of the three linear regression models and is empty for the column #Actual. Similarly, the last three columns show the predicted ship deficiency numbers, mean squared errors (MSE), and ship selection decisions based on the predictions of different linear regression models.

We can observe that LR1 with the smallest MSE among the three models has the best prediction performance. Nevertheless, if we consider the decision of the ship selection optimization model based on the predicted ship deficiency number, we can find that LR2 can induce the best decision when one ship is selected, while LR3 performs the best when two ships are selected. In contrast, the optimization model using predicted deficiency number of LR1 as the input cannot provide a satisfactory decision in either situation

because the predicted deficiency number of ship 1 is smaller than ship 2 and ship 3 by LR1. This is a counter example of our intuition that more accurate prediction of ship deficiency number supports better decisions in the following optimization model using the prediction as the input. Therefore, it can be seen that the accuracy of ship ranking, rather than the absolute ship deficiency number, is more suitable to be used to evaluate the prediction model to derive informed decisions because the optimization model is concerned with selecting the ships with a relatively larger predicted number of deficiencies.

Table 3. A counter example in the ship selection problem.

	#Actual	# Predict by LR1	# Predict by LR2	# Predict by LR3
Ship 1	10	8	20	5
Ship 2	7	10	2	16
Ship 3	5	9	6	1
MSE	/	9.67	42	16.67
Inspecting one ship	Ship 1	Ship 2	Ship 1	Ship 2
Inspecting two ships	Ship 1, 2	Ship 2, 3	Ship 1, 3	Ship 1, 2

Motivated by the above observation, we try to develop a new linear regression model with the pairwise-comparison target as the loss function, which is a target gained by comparing the ranking of ship pairs based on their deficiency numbers. In this target, we aim to minimize the prediction error of the relative ranking of each ship pair. To be specific, we hope that, if the actual deficiency number of ship i is larger than ship j , the predicted result \hat{n}_i should also be larger than \hat{n}_j . To demonstrate this relationship, we define that, for each ship $i \in \{1, \dots, N\}$, set Φ_i contains ships whose actual deficiency numbers are smaller than ship i , i.e., $\Phi_i = \{j \in \{1, \dots, N\} : n_j < n_i\}$. Then, we introduce a new variable z_{ij} to capture the prediction accuracy of the relative relationship between ship i and ship j regarding their numbers of deficiencies. If the predicted deficiency number ranking of the two ships is correct, z_{ij} is 0; otherwise, it is equal to $1 - (\hat{n}_i - \hat{n}_j)$. Notably, only when the predicted deficiency number \hat{n}_i is larger than \hat{n}_j by more than 1 would we say that the predicted deficiency number ranking of the two ships is correct.

We employ the linear regression model in Section 4.1.1 and the pairwise-comparison target described above to build a linear programming model. Moreover, to restrict the values of coefficients to a rational range, we can obtain the confidence interval of coefficients based on the linear regression model introduced in Section 4.1.1. Specifically, we set w to be in the interval $[w_{LR}^* - 3\sigma_w, w_{LR}^* + 3\sigma_w]$, where w_{LR}^* and σ_w are the best fitting and standard deviation of w obtained by the least squares method, respectively. Similarly, we can set the confidence interval of b as $[b_{LR}^* - 3\sigma_b, b_{LR}^* + 3\sigma_b]$. The coefficients w and b obtained by the least squares method are distributed as $N_{r+1}((w_{LR}^*, b_{LR}^*)^T, \sigma_n^2(X'X)^{-1})$, which is an $r + 1$ dimensional normal distribution. r is the total number of features, i.e., 14, X is the design matrix, and σ_n is the standard error of the ship deficiency number on the training set, which can be calculated by $\sum_{i=1}^N (\hat{n}_i - n_i) / (N - r - 1)$ [31].

To summarize, the linear programming model can be written as

$$\min_{z, w, b} \sum_{i=1}^N \sum_{j \in \Phi_i} z_{ij} \quad (6)$$

$$\text{s.t. } z_{ij} \geq 0 \quad \forall i \in \{1, \dots, N\}, j \in \Phi_i, \quad (7)$$

$$z_{ij} \geq 1 - (w^T x_i - w^T x_j) \quad \forall i \in \{1, \dots, N\}, j \in \Phi_i, \quad (8)$$

$$w_{LR}^* - 3\sigma_w \leq w \leq w_{LR}^* + 3\sigma_w, \quad (9)$$

$$b_{LR}^* - 3\sigma_b \leq b \leq b_{LR}^* + 3\sigma_b. \quad (10)$$

Objective (6) minimizes the sum of ranking error of all pairs of ship i and ship j , where $n_i > n_j$. Constraints (7) and (8) specify that ranking error z_{ij} is 0 if the ranking of the predicted deficiency numbers of i and j is correct; otherwise, the larger the prediction ranking error is, the larger z_{ij} is. Constraints (9) and (10) define the domains of variables w and b , respectively. This linear programming model provides the best coefficients w_{LP}^* and b_{LP}^* with the minimal prediction ranking error. Then, the deficiency number \hat{n}_0 of a new test sample with feature vector x_0 can be predicted to $(w_{LP}^*)^T x_0 + b_{LP}^*$.

4.1.3. M3: SVM Model with Pairwise-Comparison Loss Function

We can also develop an SVM model based on the pairwise-comparison target described in Section 4.1.2. SVM is a supervised machine learning model widely used for classification, whose decision boundary is the maximal-margin hyperplane learned from the training samples. In this study, we adopt the SVM model as a linear classifier for binary classification. To be specific, given a set of training samples, whose outputs are different classes, the SVM maps the feature vectors of training samples to points in space. When the training data are linearly separable, we can select two parallel hyperplanes that separate the two classes of points so that the distance between the closest points from the two classes is maximized. The maximal distance is called the optimal margin and the closest points are lying on the parallel hyperplanes, which are also boundaries of the margin. When the training data are not linearly separable, which is the case in our study, we try to maximize the margin while giving punishment if the points fall on the wrong side of the margin. The hyperparameter C determines the trade-off between a larger margin and smaller predicted errors, and we will explain the detailed meaning of C in the following paragraphs. Then, given a new sample whose target is to be predicted, the feature vector is mapped to a point in the same space and the predicted class depends on the side of the margin it falls.

Firstly, we transform the inspection records into samples in binary class based on the pairwise-comparison method. To be specific, we assume that the deficiency number of each ship $u \in 1, \dots, N$ in the training set is linear to its feature vector x_u , i.e., $\hat{n}_u = w^T x_u + b$. For a pair of ship u and ship v in the training set, we have $\hat{n}_u - \hat{n}_v = w^T x_u - w^T x_v = w^T (x_u - x_v)$ in opposite targets. Then, for each pair of ships u and v , we can generate two samples with feature vectors $x_u - x_v$ and $x_v - x_u$, respectively. For example, if $n_u \geq n_v$, we can construct a positive sample point $x_u - x_v$ with label 1 and a negative sample point $x_v - x_u$ with label -1 . We denote a new sample by k , whose input x_k is the difference of two ship feature vectors, i.e., $x_u - x_v$, and output y_k is the relative ranking of the two ships based on their deficiency number, i.e., 1 if $n_u \geq n_v$ and -1 , otherwise. Therefore, the original training data set with N records can be transformed to a new training set with K samples, where K is equal to $N \cdot (N - 1)$ because we compare the deficiency number of each pair of ships from the original data set and generate two samples from each pair to form the new data set. Our original data set contains 2643 training records, so there will be a total of $2643 \times 2642 = 6,982,806$ training records if we compare each pair of ships, which are too many to compute. Therefore, we adopt a sampling method and randomly generate 30,000 training records. Following the same procedure, we randomly generate 10,000 validation and test records, respectively.

As described in Section 4.1.2, we hope that the predicted ranking of a ship pair based on the deficiency number is the same as the actual ranking. Thus, we can define the loss of each training sample as the ranking error, which is equal to 0 if the relative ranking of two ships based on deficiency number is correct and is equal to the distance of the output point from the decision boundary, otherwise. Mathematically, the loss function of sample k in the training set is

$$l(f(x_k), y_k) = \begin{cases} \max(-f(x_k) + o, 0) & y_k = 1, \\ \max(f(x_k) + o, 0) & y_k = -1, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Notably, σ is usually a small constant set as 1, which is the threshold value for correct classification and is equal to the width of the margin in SVM. Function (11) can be demonstrated by the classic hinge loss function of the SVM model $L(f)$ as follows [32]:

$$l(f(x_k), y_k) = \max(0, 1 - y_k f(x_k)), \quad (12)$$

$$L(f) = \sum_{k=1}^K l(f(x_k), y_k) + \frac{\|w\|_2}{2C}, \quad (13)$$

where $f(x_k) = w^T x_k + b$ is the classification function and $\|w\|_2 / (2C)$ is the regularization term. Hyperparameter $C > 0$ determines the trade-off between increasing the margin size and ensuring that x_k lies on the correct side of the margin. Specifically, when C becomes smaller, the penalty for misclassification reduces and generalization ability is expected to be stronger.

Then, we can follow the standard process of SVM and train the model with a gradient descent to find the best coefficients w_{SVM}^* with loss function (13). Given a new ship with feature vector x_0 , its deficiency number \hat{n}_0 can be predicted by $(w_{SVM}^*)^T x_0$. Notably, although this predicted deficiency number may vary from the actual deficiency number by a constant term b , it will not influence the ship selection decision since all the ship deficiency numbers minus or plus a constant term will not change the relative ranking of ships.

4.2. Optimization Model

In the optimization model, we treat each unknown parameter n_m as its predicted value \hat{n}_m . Therefore, the ship selection problem can be written in a deterministic manner as follows:

$$\max_u \sum_{m=1}^M \hat{n}_m u_m \quad (14)$$

subject to constraints (2) and (3).

5. Numerical Experiments

We conduct a case study using the data described in Section 3.2 to validate the proposed three ship deficiency number prediction models M1, M2, and M3. Particularly, we first validate the three proposed prediction models in Section 5.1. Then, we compare the performances of the ship selection optimization models, which are based on the predicted ship deficiency numbers given by different prediction models in Section 5.2.

The experiments are conducted on a computer with Intel Core i5 CPU and 8 GB memory under the Mac operating system. The models are implemented in Python programming language using Gurobi 9.1.2 as the solver.

5.1. Construction of the Three Prediction Models

As described in Section 4, we first construct the three prediction models. Specifically, M1 is a linear regression model based on least squares method, in which the input is the feature vector of a ship and the output is the ship deficiency number. M2 is a linear programming model, which aims to calculate the best coefficients w_{LP}^* and b_{LP}^* that minimize the pairwise-comparison target. There is no hyperparameter in M1 or M2. M3 is an SVM model, where the input is the difference between the feature vector of two ships and the output is the relative ranking of the two ships based on their deficiency number. There is a critical hyperparameter C in SVM that determines the accuracy of the classifier. Theoretically, when the value of C becomes smaller, the training error decreases, but the model might be overfitting and the generalization ability might be weakened. When C goes to infinity, classification errors are not allowed, which is a hard-margin SVM problem; when C goes to 0, we no longer care about whether the classification is correct and only

requires a larger margin, then we may not obtain a meaningful solution, and the algorithm may not even converge. Since our training set has a total of 30,000 records, the classification error can be large and the hyperparameter C should be set as about $1/30,000$ in order to effectively trade-off between larger margin size and smaller classification error. Therefore, we change the value of C from $1/30,000$ to $1/100$ as shown in the first column of Table 4 and present the accuracy scores and computation time of SVM with different C on the training and validation sets in Table 4, where the accuracy scores are the percentages of correct predicted classifications on the whole data set.

Table 4. Prediction performances of SVM with different values of C on the training and validation sets.

C	Training Accuracy Score	Validation Accuracy Score	Computation Time (s)
$1/30,000$	0.71	0.67	0.03
$1/10,000$	0.72	0.66	0.03
$1/3000$	0.72	0.66	0.03
$1/1000$	0.73	0.66	0.04
$1/100$	/*	/	/

*: This indicates that the SVM with $C = 1/100$ cannot converge.

We can find that the accuracy score on the training set increases when C becomes larger because the punishment of classification errors increases. The SVM with $C = 1/30,000$ has the highest accuracy score 0.67 on the validation set, and the validation accuracy score stays at 0.66 when the value of C increases, which means that the SVM with $C = 1/30,000$ can correctly predict the relative rank of about 67% ship pairs in the validation set. The computation time also increases slightly as the value of C increases, while the SVM with $C = 1/100$ cannot converge. Therefore, we choose the optimal $C = 1/30,000$ as it leads to the highest accuracy score and least computation time.

Following the procedures introduced in Section 4, we train the three models respectively and present the coefficients of each model in Table 5. Moreover, the computation time and performances of the three models are shown in Table 6. Notably, M3 is trained with the optimal hyperparameter $C = 1/30,000$ gained from the above experiment. The input data of the three proposed prediction models have been standardized to eliminate the impact of orders of magnitude of different features.

Table 5. Coefficients value of w and b in M1, M2, and M3.

Model	w	b
M1	(0.62, 1.20, −1.42, −0.51, −0.23, −0.12, −0.08, −0.29, −0.05, 2.46, 1.00, 0.25, 0.12, 0.08, 0.09, 0.26, −0.39, 0, −0.29)	4.8
M2	(0.26, 0.37, −0.68, −0.25, −0.01, −0.06, 0.15, −0.09, 0, 2.14, 0.64, −0.05, 0.05, 0.02, 0.02, −0.02, −0.15, 0, −0.2)	4.54
M3	(0.03, −0.09, −0.11, −0.15, −0.13, −0.01, −0.03, −0.1, −0.12, 0.1, 0.1, 0.03, −0.04, 0.08, −0.04, 0.12, −0.02, 0, −0.1)	0

Table 6. Prediction performances of M1, M2, and M3.

Model	Metric	Training Set	Validation Set	Test Set
M1	Computation time (s)	0	/	/
	MSE	21.16	25.21	19.86
M2	Computation time (s)	1196.81	/	/
	Objective value	2,401,786.76	283,179.75	264,786.51
M3	Computation time (s)	0.11	/	/
	Accuracy score	0.71	0.67	0.64

From Table 5, we can observe that the proposed three prediction models yield different values of w and b . Particularly, the value of intercept b in M3 is 0 because our training data samples are generated symmetrical, i.e., whenever there is a positive sample point $x_u - x_v$ with label 1, there is also a negative sample point $x_v - x_u$ with label -1 . From Table 6, we can find that the computation time of both M1 and M3 is small, while the computation

time of M2 is much larger at about 1196.81 seconds. This is because that M2 with millions of variables and constraints is more difficult to solve. The MSEs of M1 on the training, validation, and test sets are similar and are 21.16, 25.21, and 19.86, respectively. The objective value of M2 on the training set is about 10 times of that on the validation and test sets because the number of records on the training set is much larger and the objective value, which demonstrates the predicted classification error, is also larger. For M3, the prediction accuracy scores on the validation and test sets are slightly smaller than that on the training set. We can find that the MSEs of M1 are small, while the objective values of M2 are large and the accuracy scores of M3 are not very high, which indicates that there may be some mis-classified cases in M2 and M3.

5.2. Decision Quality of the Three Prediction Models

To compare the decision qualities of the three ship deficiency number prediction models, we randomly select M ships from the test set as an instance set of foreign visiting ships to the port on one working day. Specifically, we assume $M = 10, 20, 30$, respectively, and 10 instance sets are generated for each M to eliminate the influence of random factors and to ensure the generality of the results. We set the number of selected ships for inspection, i.e., S , to 5. Then, we test the ship selection optimization model (14) based on the predicted ship deficiency number by M1, M2, and M3 using these instances.

We also compare the total number of deficiencies detected based on the three proposed prediction models with that based on the random selection policy and the perfect-forecast policy. In the random selection policy, we randomly select S ships from the total of M ships and repeat the random selection process for 200 times, then we calculate the mean of the total number of deficiencies detected in these 200 times as the lower bound benchmark for the ship selection problem. The perfect-forecast policy shows the best ship selection decision where the deficiency number of each ship can be accurately predicted in advance. Although the perfect-forecast policy is an ideal situation that cannot be achieved in practice, we can use the value of its objective function as an upper bound in theory to compare the performance of the three proposed prediction models. The comparison results are shown in Table 7.

Table 7. Performances of the optimization model based on the prediction results of M1, M2, and M3.

Performance	$M = 10$			$M = 20$			$M = 30$		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
Average number of detected deficiencies	25.8	25.7	28.1	38.5	38.1	42	68.8	69.6	73.5
Average improvement from random selection policy	25.37%	23.62%	40.38%	75.85%	74.59%	92.41%	179.59%	181.65%	198.55%
Average gap from perfect-forecast policy	25.18%	26.31%	15.72%	28.17%	29.02%	21.44%	19.00%	18.42%	13.05%
Max gap from perfect-forecast policy	52.38%	52.38%	42.86%	62.30%	62.30%	52.46%	32.67%	36.76%	18.81%
Min gap from perfect-forecast policy	3.70%	0.00%	3.33%	6.35%	9.84%	0.00%	11.21%	5.83%	7.32%

From Table 7, we can observe that the average number of detected deficiencies of optimization models based on all of the three proposed prediction models are larger as the ship number M increases because there are more candidate ships to select from. In particular, the optimization decision based on M3 can always detect more deficiencies whatever the value of M , while the optimization decision based on M2 detects the least deficiencies when $M = 10$ and $M = 20$, and the optimization decision based on M1 detects the least deficiencies when $M = 30$. Taking the random selection policy as a benchmark, the results verify our intuition that the performances of the optimization models based on the three prediction models are always better than the random selection policy. Moreover,

the average improvements of the random selection policy increase sharply from about 20% to about 170% as M becomes larger. This is because, when the value of M becomes larger, there are more ships to choose from and the number of ships with larger deficiency numbers also increases. Our proposed ship selection methods can effectively target those ships with larger ship deficiency numbers, which results in larger improvements over the random selection policy. M3 stands out among the three prediction models if compared with the perfect-forecast policy as it has smaller average gaps from the perfect-forecast policy, i.e., 15.72% when $M = 10$, 21.44% when $M = 20$, and 13.05% when $M = 30$. Moreover, M3 has significant improvement over the random selection policy, which are 40.38% when $M = 10$, 92.41% when $M = 20$, and 198.55% when $M = 30$, showing its high efficiency in ship selection. The optimization model based on M2 performs worst among the three prediction models when $M = 10$ and $M = 20$, but it performs better than M1 when the value of M increases to 30. In addition, the performance of the optimization model based on M2 is not stable because the minimum gap of the optimization model based on M2 from the perfect-forecast policy is the smallest when $M = 10$ and $M = 20$, but its maximum gaps are also larger or equal to the other two models. By contrast, optimization models based on M3 always provide decisions with smaller maximum and minimum gaps from the perfect-forecast policy than M1. Notably, while the optimization model based on M2 and M3 can make the best decision in some cases, i.e., the minimum gap from the perfect-forecast policy is 0, M1 cannot make the same decision as the perfect-forecast policy on any instance, which shows its limitation.

To summarize, from the computation time perspective, we can observe that M1 and M3 are more efficient; from a model performance perspective, the optimization model based on M3 can provide more stable and effective solutions with smaller average gaps from the perfect-forecast policy whatever the value of M is. Moreover, the optimization model based on M2 can provide better decisions than that based on M1 when the value of M is larger, i.e., $M = 30$.

6. Conclusions

PSC inspection is an essential method to protect maritime safety, the marine environment, and the rights of seafarers. However, inspecting all foreign visiting ships is neither practical, due to limited inspection resources at the port, nor necessary, as only a small fraction of ships are actually substandard. Therefore, to improve the effectiveness of ship inspection, how to select high-risk ships for inspection is one critical issue faced by MoUs on PSC. In this study, we construct three prediction models, namely M1, M2, and M3, to estimate the deficiency number of individual visiting ships, based on which we further develop a ship selection optimization model for the port to select ships with a higher risk for inspection. Specifically, M1 is a linear regression model with a classic loss function, i.e., MSE of deficiency number, while M2 and M3 are linear prediction models with the pairwise-comparison target, which is a target motivated by the property of the following optimization that aims to minimize the ranking errors of ship pairs based on their deficiency numbers. We achieve this pairwise-comparison target by a linear programming model in M2 and an SVM model in M3.

To validate our proposed models, we use the Hong Kong port as a case study to construct the three prediction models and to evaluate the performances of the optimization models based on each prediction model. We first train M3 with different values of hyperparameter C and find that the SVM model with $C = 1/30000$ performs best. Results of prediction experiments show that M1 and M3 have smaller computation time, while M2 requires more time since it has millions of variables and constraints. By comparing the performances of the optimization model based on the prediction results of M1, M2, and M3 with the random selection policy, we can observe that the improvements of all of the three prediction models increase sharply from about 20% to about 170% when the value of M increases from 10 to 30, which indicates that an optimization model based on a proper prediction of ship deficiency number can perform much better than random policy.

In addition, the optimization model based on M3 always has larger average improvements from random selection policy and smaller average gaps from perfect-forecast policy among the three models, while M2 performs worst on both indicators except when $M = 30$. We can also observe that M2 and M3 can provide the same decision as the perfect-forecast policy, while M1 fails to make the best decision in any case.

The main contributions of this paper lie in the following aspects:

1. Practically, we concentrate on the ship selection problem, which is essential for the port state to identify ship noncompliance (i.e., deficiencies) with limited inspection resources. Our proposed semi-SPO method can inspire the port states to determine the ship inspection priority more effectively by predicting the relative ranking of ships based on their risk rather than the absolute risk value of each ship, which is the determinant to generate efficient ship inspection decisions. Moreover, the semi-SPO ship inspection scheme can be applied in the maritime industry to target high-risk ships efficiently without too much computation time. Therefore, the effectiveness of the current ship inspection planning process in practice can be improved, and the main objectives of PSC to eliminate substandard shipping and safeguard the sea can be enhanced.
2. Theoretically, we implement the pairwise-comparison based semi-SPO method for ship inspection planning, which partially combines prediction and optimization models by considering the structure and property of the optimization model when constructing the machine learning model. We also compare our semi-SPO method with the widely adopted prediction model, which treats prediction and optimization models as sequential steps.
3. The numerical experiments show that, while the prediction performance of M2 and M3 is not obviously better than M1, the optimization model based on the results of M3 is much better than that of M1 within the similar computation time. In addition, our semi-SPO method, i.e., M2 and M3, can provide the best decision as the perfect-forecast policy in some cases, while M1 cannot make the perfect decision in any instance. The experiments reveal the effectiveness and stability of the semi-SPO method, especially M3, for ship inspection planning.

Future research could be conducted from the following aspects: first, whereas the optimization model based on M3 can provide a better decision than that of M1 by taking the structure and properties of the optimization model into consideration, the prediction accuracy can be further improved by adopting other prediction models such as the Bayesian network or tree-based model; second, M2 based on the semi-SPO method cannot perform well as we expected. This may be due to its limitation on the construction of its objective, which only takes the classification errors into account and the generalization ability is weakened. Therefore, other objective construction could be explored in the future.

Author Contributions: Conceptualization, H.W.; methodology, Y.Y., R.Y. and H.W.; software, Y.Y.; validation, Y.Y.; formal analysis, Y.Y.; investigation, Y.Y., R.Y. and H.W.; resources, R.Y.; data curation, Y.Y. and R.Y.; writing—original draft preparation, Y.Y. and R.Y.; writing—review and editing, Y.Y. and R.Y.; visualization, Y.Y.; supervision, H.W.; project administration, H.W.; funding acquisition, H.W. and R.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Guangdong Basic and Applied Basic Research Foundation Grant No. 2019A1515011297.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in the Asia Pacific Computerized Information System at <https://apcis.tmo.org/> (accessed on 1 December 2021) and World Register of Ships database http://www.tokyo-mou.org/inspections_detentions/psc_database.php (accessed on 1 December 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yu, M.; Fransoo, J.C.; Lee, C.Y. Detention decisions for empty containers in the hinterland transportation system. *Transp. Res. Part B Methodol.* **2018**, *110*, 188–208. [CrossRef]
2. Luo, M.; Shin, S.H. Half-century research developments in maritime accidents: Future directions. *Accid. Anal. Prev.* **2019**, *123*, 448–460. [CrossRef] [PubMed]
3. Bell, M.G.; Pan, J.J.; Teye, C.; Cheung, K.F.; Perera, S. An entropy maximizing approach to the ferry network design problem. *Transp. Res. Part B Methodol.* **2020**, *132*, 15–28. [CrossRef]
4. IMO. Initial IMO GHG Strategy. 2019. Available online: <https://www.imo.org/en/MediaCentre/HotTopics/Pages/Reducing-greenhouse-gas-emissions-from-ships.aspx> (accessed on 25 July 2022).
5. IMO. Maritime Safety. 2019. Available online: <https://www.imo.org/en/OurWork/Safety/Pages/default.aspx> (accessed on 25 July 2022).
6. Zhang, P.; Zhao, L.; Vata, O.; Rajagopal, S. Restructuring seafarers' welfare under the Maritime Labour Convention: An empirical case study of Greece. *Marit. Bus. Rev.* **2020**, *5*, 373–389. [CrossRef]
7. IMO. Resolution A. 1155(32) Adopted on 15 December 2021 (Agenda Items 12 and 14) Procedures for Port State Control, 2021. 2022. Available online: <https://www.register-iri.com/wp-content/uploads/A.115532.pdf> (accessed on 2 February 2022).
8. Wang, S.; Yan, R.; Qu, X. Development of a non-parametric classifier: Effective identification, algorithm, and applications in port state control for maritime transportation. *Transp. Res. Part B Methodol.* **2019**, *128*, 129–157. [CrossRef]
9. Heij, C.; Bijwaard, G.E.; Knapp, S. Ship inspection strategies: Effects on maritime safety and environmental protection. *Transp. Res. Part D Transp. Environ.* **2011**, *16*, 42–48. [CrossRef]
10. Tokyo MoU. Information Sheet of the New Inspection Regime (NIR). 2014. Available online: <http://www.tokyo-mou.org/doc/NIR-information%20sheet-r.pdf> (accessed on 4 July 2022).
11. Yan, R.; Wang, S.; Peng, C. Ship selection in port state control: Status and perspectives. *Marit. Policy Manag.* **2022**, *49*, 600–615. [CrossRef]
12. Yang, Z.; Yang, Z.; Yin, J. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. *Transp. Res. Part A Policy Pract.* **2018**, *110*, 38–56. [CrossRef]
13. Elmachet, A.N.; Grigas, P. Smart “predict, then optimize”. *Manag. Sci.* **2022**, *68*, 9–26. [CrossRef]
14. Demirović, E.; Stuckey, P.J.; Bailey, J.; Chan, J.; Leckie, C.; Ramamohanarao, K.; Guns, T. An investigation into prediction+ optimisation for the knapsack problem. In Proceedings of the International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research, Thessaloniki, Greece, 4–7 June 2019; pp. 241–257.
15. Yan, R.; Wang, S. Ship inspection by port state control—Review of current research. In *Smart Transportation Systems 2019*; Springer: Singapore, 2019; pp. 233–241.
16. Li, K. The safety and quality of open registers and a new approach for classifying risky ships. *Transp. Res. Part E Logist. Transp. Rev.* **1999**, *35*, 135–143. [CrossRef]
17. Degré, T. The use of risk concept to characterize and select high risk vessels for ship inspections. *WMU J. Marit. Aff.* **2007**, *6*, 37–49. [CrossRef]
18. Xu, R.F.; Lu, Q.; Li, W.J.; Li, K.; Zheng, H.S. A risk assessment system for improving port state control inspection. In Proceedings of the 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, China, 19–22 August 2007; Volume 2, pp. 818–823.
19. Xu, R.; Lu, Q.; Li, K.; Li, W. Web mining for improving risk assessment in port state control inspection. In Proceedings of the 2007 International Conference on Natural Language Processing and Knowledge Engineering, Beijing, China, 30 August–1 September 2007; pp. 427–434.
20. Gao, Z.; Lu, G.; Liu, M.; Cui, M. A novel risk assessment system for port state control inspection. In Proceedings of the 2008 IEEE International Conference on Intelligence and Security Informatics, Taipei, Taiwan, 17–20 June 2008; pp. 242–244.
21. Chi, Z.; Jun, S. Automatically optimized and self-evolutional ship targeting system for port state control. In Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 791–795.
22. Yang, Z.; Yang, Z.; Teixeira, A.P. Comparative analysis of the impact of new inspection regime on port state control inspection. *Transp. Policy* **2020**, *92*, 65–80. [CrossRef]
23. Yang, Z.; Yang, Z.; Yin, J.; Qu, Z. A risk-based game model for rational inspections in port state control. *Transp. Res. Part E Logist. Transp. Rev.* **2018**, *118*, 477–495. [CrossRef]
24. Dinis, D.; Teixeira, A.; Soares, C.G. Probabilistic approach for characterising the static risk of ships using Bayesian networks. *Reliab. Eng. Syst. Saf.* **2020**, *203*, 107073. [CrossRef]
25. Yan, R.; Wang, S.; Peng, C. An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities. *J. Comput. Sci.* **2021**, *48*, 101257. [CrossRef]
26. Heij, C.; Knapp, S. Shipping inspections, detentions, and incidents: An empirical analysis of risk dimensions. *Marit. Policy Manag.* **2019**, *46*, 866–883. [CrossRef]
27. Knapp, S.; Heij, C. Improved strategies for the maritime industry to target vessels for inspection and to select inspection priority areas. *Safety* **2020**, *6*, 18. [CrossRef]
28. Tokyo MoU. Annual Report 2016 on Port State Control in the Asia-Pacific Region. 2017. Available online: <http://www.tokyo-mou.org/doc/ANN16.pdf> (accessed on 4 July 2022).

29. Tokyo MoU. Black–Grey–White Lists. 2017. Available online: <http://www.tokyo-mou.org/doc/Flag%20performance%20list%202017.pdf> (accessed on 4 July 2022).
30. Paris MoU. Criteria for Responsibility Assessment of Recognized Organizations (RO). 2013. Available online: https://www.parismou.org/sites/default/files/RO%20responsibility%20rev11_0.pdf (accessed on 4 July 2022).
31. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 2002; Volume 5.
32. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2.