Evaluation and Prediction of Punctuality of Vessel Arrival at Port: A Case Study of Hong Kong

Zhong Chu, Ran Yan *, and Shuaian Wang

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abstract

The punctuality of vessel arrival at port is a crucial issue in contemporary port operations. Although vessels are usually required to report their estimated time of arrival (ETA) on the way to the destination port, vessels' actual time of arrival (ATA) is generally different from the reported ETA as there are several factors (including unexpected and rough weather and sea conditions, unexpected operational inefficiency, and unexpected port congestion) cause their ATA to differ from their reported ETA. Uncertainties in vessel arrival may lead to port handling inefficiency, resulting in economic losses. Therefore, evaluating, predicting, and then optimizing vessel arrival time at a port can improve terminal operational efficiency and optimize port resource allocation. We first analyze ship arrival punctuality and predict ship arrival times using vessel visiting data in 2021 for the Hong Kong Port (HKP). We also quantitatively evaluate vessel arrival uncertainty in different time slides prior to arrival at the port. Our results confirm that the overall vessel arrival uncertainty decreases as vessels approach the HKP. In this paper's prediction section, we implement a machine learning approach to predicting vessel arrival time based on vessels' historical arrival data and vessel generic features. Our prediction model can reduce the error in the prediction of ship ATA data by approximately 40% (from 25.5h to 15.5h) using the root mean squared error metric and 20% (from 13.8h to 11.0h) using the mean absolute error metric compared with the reported ETA data. The proposed vessel arrival time evaluation and prediction models are applicable to port management and operation, and they can lay the foundation for future research on optimizing ports' daily operations.

Keywords- Maritime transport; Vessel arrival punctuality; Vessel arrival prediction; Port management

1 Introduction

Recent decades have witnessed dynamic developments in international shipping. According to the 2019 report of the United Nations Conference on Trade and Development, shipping is a backbone of global trade and the international supply chain, as approximately 90% of cargo is carried by vessels [33, 12, 11, 22]. Port operations planning is carried out every day with the aim of maximizing the efficiency of vessel handling operations, and the estimated time of arrival (ETA) records reported by vessels are an important reference [38]. In daily port operations, one of the crucial challenges faced by port operators is the uncertainty of vessel arrival [17], i.e., inaccuracy in ETA records. Vessels are usually required to report their ETA data before arriving at the port, but those data sometimes largely differ from the corresponding actual time of arrival (ATA), as several factors (e.g., unexpected and rough weather and sea conditions, unexpected operational inefficiency, and unexpected port congestion) cause the deviation of ETA and ATA and sequentially affect port operational efficiency [14, 17, 38]. According to the Drewry analysis of container service reliability, in November 2015, global container services only reached an average on-time performance rate of 73% if the delay threshold was set at 24 hours [8]. To mitigate the uncertainty of vessel arrival, quantitative methods can be implemented to estimate the deviation of original ETA records and obtain more accurate ETA records, which can further the decision-making process in port operations [38, 17].

*Corresponding author: angel-ran.yan@connect.polyu.hk

We aim to analyze and predict vessel arrival delays at the Hong Kong Port (HKP). The HKP is one of the busiest ports in the world, handling approximately 18 million 20-foot equivalent units (TEUs) of containers in 2020 [14]. The Hong Kong Marine Department requires every vessel planning to visit the HKP to report its ETA every 20 minutes during the 36 hours before arrival, and the port automatically records each vessel's ATA data once it arrives [6]. Based on these ship arrival data, we explore the punctuality of ship arrival at the HKP in different time slices ahead of one visit and predict vessel arrival time at the HKP using a machine learning model. For the dataset, 2,943,388 ETA data records and 13,692 ATA data records for the HKP from January 1, 2021, to December 31, 2021, are collected. We find that the overall time difference between ETA and ATA decreases as vessels approach the port. Next, a random forest (RF) model is developed to predict vessels' arrival time using vessel history arrival data and vessel generic features (e.g., length, beam, and gross tonnage (GT)). The results show that our prediction model can reduce 40% of the deviation error evaluated by the root mean square error (RMSE), from 25.5 hours to 15.5 hours, and 20% of the error evaluated by the mean absolute error (MAE), from 13.8 hours to 10.9 hours, on the reported ETA data. Furthermore, we discuss the insights and extension of the prediction model and our results from the following three perspectives: port operations, rational commercial decisions, and advisable policy proposals. In addition, we explore and discuss several further research questions based on this study.

The rest of this paper is organized as follows. We carry out a comprehensive literature review on the evaluation and prediction of vessel arrival delay in Section 2. In Section 3, we briefly introduce the HKP and the vessel arrival dataset. Moreover, we conduct an all-inclusive statistic analysis of ship arrival data at the HKP. In Section 4, we evaluate and analyze the punctuality of vessel arrival at the HKP. In Section 5, an RF model is built to predict vessel arrival time at the HKP with vessel historical arrival data and vessel generic features. The analysis, extension, and insights of the prediction results are presented in the same section. Our final conclusion is discussed in Section 6.

2 Literature review

A considerable number of studies address the issue of vessel arrival uncertainty in ports and predict ship arrival time to assist port operators in making decisions [16, 18, 3, 38]. Most of these studies use ship sailing information extracted from automatic identification systems (AIS) as the dataset. Pursuant to International Maritime Organization (IMO) regulations, vessels of more than 500 GT are required to install AIS to avoid collisions [37, 20]. With AIS, vessel static (name, size, MMSI, and IMO number), dynamic (speed, location, and heading degree) and voyage-related (destination port, draft) records can be generated and reported every few minutes. AIS provides a powerful information-rich vessel movement dataset for researchers to conduct related studies. Starting with AIS data mining, several studies use AIS data to estimate and improve the accuracy of ETA data.

For example, Kim et al. (2017) incorporate a classification and regression tree (CART) model with a case-based reasoning framework to detect vessel delay using online vessel tracking data and historical AIS data [7]. Parolas et al. (2016) implement support vector machines and neural networks to predict ships' ETA in the Rotterdam port by combining vessel AIS data with global positioning satellite (GPS) data [20]. Dejan et al. (2020) use a machine learning-based system to predict vessel turnaround time and ETA data with 11 years of historical data for the port of Bordeaux [23]. Alfredo et al. (2018) tackle ETA estimation using a data-driven path-finding algorithm with historical AIS and long-range identification and tracking data [1]. Adrian et al. (2020) compare the performance of various machine learning models in predicting vessel delay tasks based on hundreds of features extracted from AIS data [27]. Takahiro et al. (2021) propose a Bayesian learning method to predict vessel route and voyage speed to port considering weather conditions with AIS data [15]. Park et al. (2021) reconstruct vessel trajectories using reinforcement learning and data mining on AIS data. They then introduce Bayesian sampling to estimate vessels' speed and arrival time [18]. Xu et al. (2022) estimate vessel arrival time based on a trajectory clustering method using AIS data [31].

Although AIS data have been widely studied and used by companies and academia in recent years, these data, and related research, have several shortcomings. For example, with respect to data accessibility, AIS data are highly regionalized and difficult for individuals to obtain [37]. Furthermore, the collection of AIS data is strictly regulated, and most AIS data are not publicly accessible [37, 23]. Additionally, AIS data sometimes feature errors and inaccuracies due to manually entered mistakes [37]. Moreover, empty or relatively imprecise dynamic and voyage-related data may be generated frequently with AIS [29]. Using data sources other than AIS, Salleh et al. (2017) implement a probabilistic model called the fuzzy rule-based Bayesian network to predict the port arrival punctuality of container ships [21]. In that study, instead of validating the model using a large dataset, the authors verify the proposed model on just two selected vessels, and the model requires dozens of vessel features as input. To evaluate the quality of vessel arrival, Chung et al. (2011) analyze and investigate the key factors that affect the reliability of liner shipping [5]. Albert et al. (2021) present several measurement methods to quantitatively evaluate and compare ETA prediction results across different models and vessels [25].

Based on previous studies, we note that as an alternative to utilizing AIS data, there are only a few studies that evaluate and predict the punctuality of vessel arrival at port from a big data perspective. Furthermore, few studies use practical data from the HKP. The studies that are the most similar to ours are those of Pani et al. (2015) and Yu et al. (2018) [38, 17]. Pani et al. (2015) adopt logistic regression, CART, and RF models to predict vessel arrival time at the Cagliari and Antwerp ports. Yu et al. (2018) implement back-propagation network, CART, and RF models to estimate the delay or advance of ship arrival and to evaluate the results of their predictions in daily port operations [38, 17]. In both studies, instead of providing exact predictions of arrival times, the authors predict only the vessel delay time interval; neither of them explore the port delay scenario. To fill this research gap, we conduct a quantitative vessel punctuality analysis and precisely predict vessel arrival time at the HKP based on a free public dataset [14, 6].

3 Data statistics analysis on the vessel arrivals

3.1 Background of the HKP

The HKP, located in the South China Sea, is a deepwater port that mainly provides services related to containerized manufactured products. It is one of the busiest and most efficient international container ports in the world [14]. The HKP handled nearly 18 million TEUs of containers in 2021, and it is the 10th largest container port in the world [14]. In June 2022, the HKP served approximately 270 international container liner services per week, connecting more than 600 destinations worldwide [14]. The main container terminals (CTs) of the HKP include five operators and 24 berths and are located in the Kwai Chung-Tsing Yi basin. In this study, we regard the terminals and berths at the HKP as a whole and explore and predict the punctuality of vessel arrival at the HKP.

3.2 Dataset description

The Hong Kong Maritime Department updates the vessel arrival and departure information of ocean-going vessels every 20 minutes on a government website [6], which is publicly available. The operator of a vessel due to arrive at the HKP may report multiple ETA data records on its way to the port, and the vessel's ATA record will be automatically generated once it arrives. Alternatively, even if a vessel reports its ETA data, it may not have the corresponding ATA data, which means that the vessel did not actually arrive at the HKP due to a temporary change in its route. The website contains four files: vessels that arrived in the last 36 hours, vessels that are due to arrive in the next 36 hours, vessels in port, and vessels that departed in the last 36 hours. The available variables and their explanations are shown in Table 1:

Table 1: Variables available and descriptions

Item	Description	Note
Vessel name	Name of the vessel	\
Ship type	Type of vessel	14 types in total
Trip status	Vessel trip status	Approved or pending
Agent name	Name of the vessel's agent	\
Flag	Vessel registration country	\
ETA	Vessel estimated arrival time	Provided by the vessel operator
ATA	Vessel actual arrival time	Recorded once vessel arrives at the HKP
ADT	Vessel actual departure time	Recorded once vessel depart from the HKP
Report time	Data upload time onto the website	Provided by the maritime department
Last port	Name of a ship's last port of call before arrival	Provided by the vessel operator
IMO number	The International Maritime Organization (IMO) number of a ship	Unique seven-digit number
Call sign	An alphanumeric code that uniquely identifies a vessel for radio communication	A unique identifier
Last berth	Vessel's last berthing location	\
Arrived location:	The first location where a ship stays after arriving in the Hong Kong waters	\

The variables in each file are listed in Table 2:

Table 2: Variables in each file

	Vessels arrived in last 36 hours	Vessels due to arrive in the next 36 hours	Vessels departed in the last 36 hours	In port vessels
Update frequency	Every 20 minutes	Every 20 minutes	Every 20 minutes	Every 20 minutes
Vessel name	✓	✓	✓	✓
Ship type		✓		✓
Trip status		✓		
Agent name	✓	✓		✓
Flag		✓		✓
ETA		✓		
ATA	✓			✓
ADT			✓	
Report time	✓	✓	✓	✓
Last port		✓		
IMO number				✓
Call sign	✓	✓	✓	✓
Last berth			✓	
Arrived location:	✓			✓

To explore vessel arrival patterns and improve the efficiency of daily operational planning at the HKP, we collect data from the website covering the period from January 1, 2021 to December 31, 2021. From the above data description, we can find that a vessel's ATA and ETA data are in different files and thus should be paired for further processing. In addition, a vessel only generates a fixed ATA record once it arrives at the port, whereas it updates its ETA data every 20 minutes during its voyage to the port. The reported ETA data records may change as the arrival time approaches and there are missing or abnormal ETA and ATA data, as the data records on the website are reported and uploaded by the vessels' operators and masters. Therefore, it is necessary to pre-process the vessel arrival data and pair a vessel's ETA with its corresponding ATA data before evaluating the punctuality of vessel arrival and predicting the actual time of vessel arrival at the HKP.

3.3 Data collecting and pre-processing

The basic steps for data collecting and pre-processing are listed as follows:

(1) Obtain ETA and ATA data:
Every 20 minutes, the system generates four files containing the arrival and departure information of ocean-going vessels on the website [6]. To match the vessels' ETA and ATA data for further research, the first step is to gather these ETA and ATA data from separate folders to form an ETA dataset and an ATA dataset.

(2) Unify the time format in the datasets:
Time formats in different datasets can be different; thus, they need to be unified. Time formats for ETA, ATA, and report time are unified to "Year-Month-Day Hour:Minute:Second" format. Records with missing time information are deleted from the dataset.

(3) Delete records with an ATA later than the report time:
Because the system will record and upload a vessel's ATA after it arrives at the HKP, data with an ATA later than the report time can be considered erroneous and are thus deleted from the dataset.

(4) Drop duplicate data from the ATA dataset:
Every 20 minutes, a file containing information about vessels that arrived in the past 36 hours is generated by the system; one vessel arrival corresponds to one ATA record. After merging the data, we find that one ship arrival may correspond to more than one ATA record with a different report time. For these ATA records, only the first record is kept; the others are deleted.

(5) Pair ETA data with the corresponding ATA data based on report time and call sign:
Every vessel has a unique call sign that is also used as the identifier in the dataset. It is also used as the identifier to sort the ETA and ATA data of a target vessel in the datasets. However, a vessel may arrive at the HKP several times in a year and for each arrival, several ETA records are generated in the system. Therefore, we cannot match ETA data with the corresponding ATA data using the call sign alone. To resolve this issue, we use the report time as a pointer to pair the data. First, we collect all records with the same call sign (i.e., from one ship) and sort their ETA and ATA data using their respective report times. Second, we compare the report time of the ETA and ATA data. For each ATA record, if the ETA report time is earlier than the ATA data, we pair them and ignore the paired ETA and ATA data in the next round of matching. For example, suppose that a dataset involves two vessels' ETA and ATA records. In the pairing stage, we first use the call sign to classify those ETA and ATA records as belonging to two ships and rank them in chronological order. Next, we pair the ATA data with their corresponding ETA data by comparing the report times. We choose the earliest ATA record and pair it with the ETA records whose report time is earlier than the selected ATA record.

(6) Delete ETA data that lack corresponding ATA data:
After conducting the pairing step, any ETA data that lack corresponding ATA data can be regarded as erroneous. These ETA data occur because a vessel reports its ETA data but does not actually arrive at the HKP. Consequently, no ATA record is generated. Accordingly, these ETA records cannot be matched with corresponding ATA records.

(7) Delete ETA data when the time difference between the ETA data and the corresponding ATA data is more than 5 days:
A vessel starts reporting ETA data 36 hours before its arrival at the port, so we choose 120 hours as the threshold for filtering ETA and ATA data that differ too much. If the time difference between ETA and ATA data is greater than the threshold, we regard the ETA data as erroneous and delete them. For example, suppose that there is a vessel that passes through Hong Kong waters without actually arriving at the HKP at the beginning of the month. Then, the vessel visits and arrives at the HKP in the middle of the month. The system records the ETA data of the two voyages, which occur at the beginning and in the middle of the month, but only one ATA record, which is created in the middle of the month. If no filter is used, all of the ETA records will be paired with the single ATA record. By considering the threshold proposed in this step, only the ETA record of the second voyage will be matched with the ATA record.

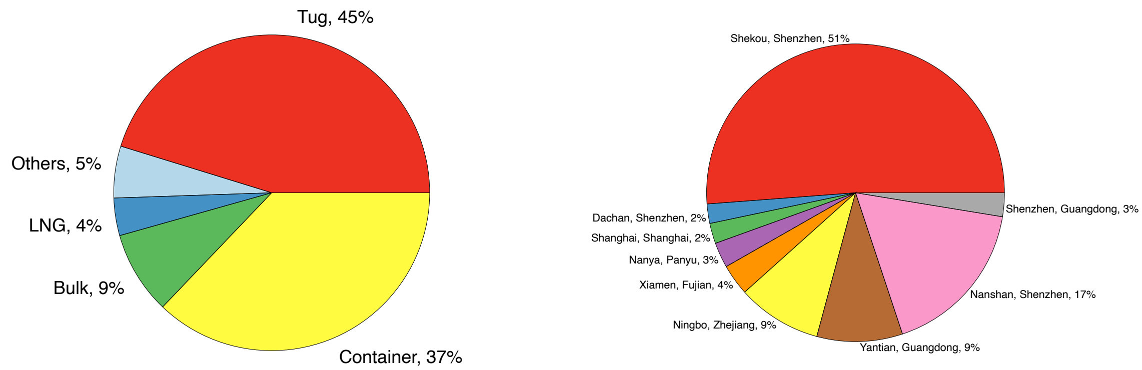
Before data pre-processing, there are 2,943,388 ETA records and 1,638,368 ATA records in the dataset. After the data cleaning and pairing processes, 1,546,443 ETA records and 13,637 ATA records remain and are paired. The ETA dataset contains 23,789 records if we drop all duplicate records regardless of report time. The processes and number of data records involved are summarized in Table 3:

Table 3: A summary of the steps involved in data pre-processing and combining

Step number	Step goal	ETA data records left	ATA data records left
1	Obtain and combine data	2,943,388	1,638,368
2	Unify time format	2,940,139	1,635,236
3	Drop duplicate ATA data	2,940,139	13,692
4	Delete unmatched ETA and ATA data	2,001,324	13,637
5	Delete mismatched ETA data	1,546,443	13,637

3.4 Statistical analysis on vessel arrival data

In this section, we focus on the following two types of vessel arrival data: paired vessel arrival data and unpaired vessel arrival data. We conduct a comprehensive statistical analysis of vessel arrival types, days, months, ETA change times, and vessel delays.



(a) Vessel type analysis in unpaired ETA data

(b) Last port of call in unpaired ETA data

Figure 1: Distribution of features in unpaired ETA data

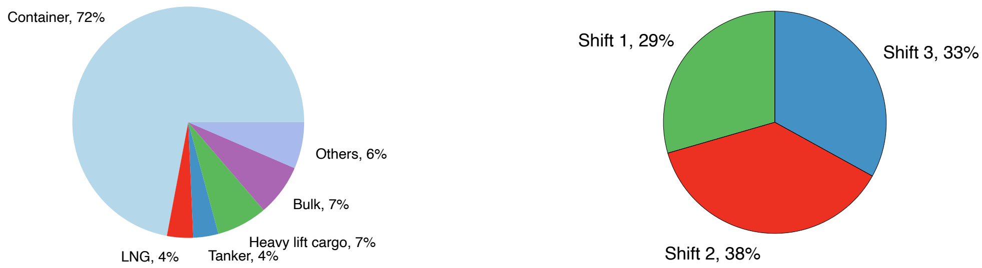
3.4.1 Statistical analysis of unpaired vessel arrival data

After data pre-processing and pairing, 34,394 ETA records and 55 ATA records are unpaired. Fifty items in the unpaired ATA data are due to time format errors, and the reason for the other unpaired data is that the vessel that arrived only updated its ATA records, not its ETA data. Of the unpaired ETA records, 24,551 have no corresponding ATA data, and 9,843 records are deleted because the time difference between ATA and ETA is more than 5 days.

Next, we analyze the unpaired ETA data from the perspective of vessel type and last port information. The visualization results are shown in Figure 1(a) and 1(b). As can be seen in Figure 1(a), the vessel types of the ships with unpaired ETA data are biased toward tugs and container ships, which correspond to approximately 80% of all records. The reason why there are many tugs with unpaired ETA data may be that when a large vessel is about to arrive at the port, several tugs are needed to assist the ship in docking, and the masters of the tugs will only report their ETA data in this process, as the tugs will not actually dock at the port. Accordingly, tugs make the greatest contribution of unpaired ETA data. The reason for the high proportion of container ships in the unpaired data is that container ships are the most common type of vessel to visit the HKP. The HKP's large base of container ships leads to their high level of representation in the proportion of ships with unpaired ETA. The results in Figure 1(b) show that for the statistics on vessels' last port of call in unpaired ETA data, the 10 most frequent ports are in mainland China and are related to 25,534 records (71.3%). Instead of long-distance trans-oceanic routes, the routes from these ports to the HKP are short-haul routes. The results also show that the vessels from these 10 ports are more likely to report misleading ETA data than vessels from other ports, which means that when a vessel passes through Hong Kong waters and reports its ETA data, it may not actually go to the HKP. The reason for the presence of these misleading ETA data in the system is that the Hong Kong Marine Department requires vessels passing through the waters near the HKP to upload their ETA data, even if they ultimately will not call at the HKP.

3.4.2 Statistical analysis on paired vessel arrival data

In this section, we analyze ship arrival where the reported ETA and ATA can be matched. First, we analyze the types of visiting ships. Fourteen types of vessels arrive at the HKP, including container, bulk, heavy lift cargo, tanker, liquefied natural gas (LNG), multi-purpose, passenger, fishing, tug, LNG tanker, nuclear fuel, car carrier, and other types of vessels. Because the top five ship types with the highest frequency constitute 93.4% of all of the visiting ships, we only keep the top five vessel types (container, bulk, heavy lift cargo, tanker, and LNG) and classify the remaining eight vessel types into the "other" category. The frequency of vessel types in the matched records is shown in Figure 2(a):



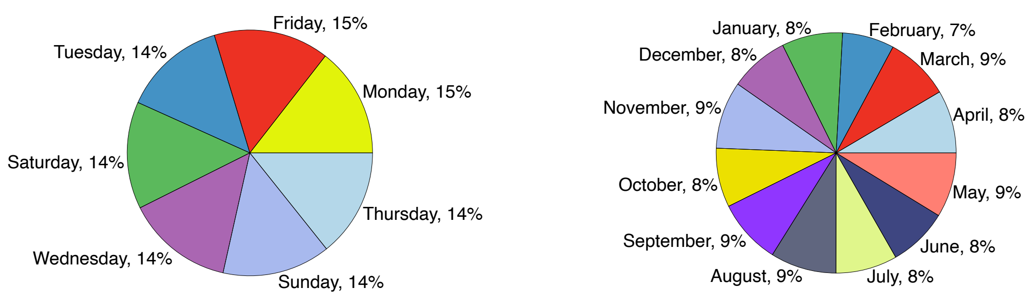
(a) Types of vessels arriving to the HKP

(b) Arrival time shifts of vessels visiting the HKP

Figure 2: Distribution of information on arriving vessels

Figure 2(a) shows that 72% of the arriving vessels are container ships, which confirms that the HKP continues to flourish as a container hub port [14]. For the other types of visiting ships, bulk carriers constitute 7.2% and heavy lift cargo vessels constitute 7.1%. In contrast, only 3.6% of the visiting ships are LNG vessels, and 3.5% of them are tankers.

Next, we analyze the ATA data using the shifts in a day, the days of the week, and the months of the year. The shifts in a day are classified into three types: Shift 1 (from 0:00 to 8:00), Shift 2 (from 8:00 to 16:00), and Shift 3 (from 16:00 to 24:00) [38]. The results of our statistical analysis are shown in Figure 2(b), Figure 3(a), and Figure 3(b), respectively.



(a) Arrival weekdays of vessels visiting the HKP

(b) Arrival months of vessels visiting the HKP

Figure 3: Distribution of time information on arriving vessels

As shown in Figures 2(b), 3(a), and 3(b), the vessels' ATA data at the HKP are slightly unevenly distributed in terms of shifts, weekdays, and months. The main conclusions from the above three figures are as follows.

(1) In terms of time shifts, most of the vessels arrive during Shift 2 (8:00∼16:00), whereas Shift 1 (0:00∼8:00) witnesses the least arrivals. The reason for this result is the higher nighttime port operating costs and crew salaries. Consequently, vessels prefer to arrive at the HKP during the day.

(2) With regard to arrival days, there is no significant difference among the days of the week regarding the number of visiting vessels. Because the HKP operates 7 days a week, 24 hours a day, vessels do not deliberately change their time of arrival. Therefore, the distribution of ship arrival days is nearly even throughout the week.

(3) February (7%) is the month of the year with the fewest arrivals. For 2021, this result is attributable to the fact that February has only 28 days, and the Spring Festival, one of Hong Kong's most important festivals, is that month. Both factors decrease the monthly number of ships visiting the HKP. In addition, December receives a low number of ships because of the Christmas holiday. Furthermore, because the HKP is in China and receives many Chinese vessels, we can anticipate that the impact of the Spring Festival is higher than that of the Christmas holiday. Therefore, February has the lowest number of arriving vessels.

Furthermore, the distributions of the three indicators are relatively constant, meaning that the ships show no significant tendency or preference related to the shift, weekday, and month of their arrival. The main reason is that most visiting vessels are operated on fixed schedules, such as liner ships [28], even during holidays such as the Chinese New Year and Christmas. Meanwhile, the shipping schedule is not necessarily a multiple of 7 days. Tramp ships without fixed schedules, such as bulk, tug, and LNG tankers, arrive at random times. Therefore, although the distributions of the arrival vessels' time shifts, weekdays, and months are not completely uniform, they are basically stable.

A vessel will report a number of ETA records on its way to the port, and its ETA data may change during its approach. Accordingly, we also analyze the change times and accuracy trend of the vessels' reported ETA data. The statistics for the vessels' ETA change times are shown in Figure 6.

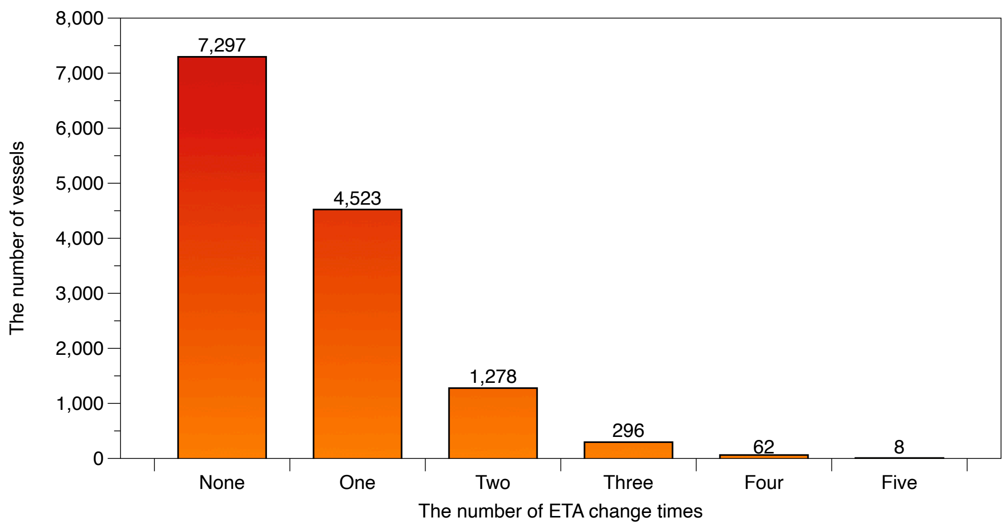


Figure 4: Vessel ETA change times

Figure 4 shows that more than half of the vessels (7,297) report only one ETA record when they are sailing to the port, and only a few vessels update their ETA data more than three times (366) during the process. Because a vessel's ETA data are reported by the ship captain when approaching the port, many factors affect the ship's actual port arrival time, which in turn affects the accuracy of the reported ETA. Typical factors include, but are not limited to, unexpected adverse weather conditions on the way and at the port, emergencies, and unexpected changes (e.g., a change in the destination). Then, ETA records theoretically become more and more accurate as the vessel approaches the port, and some captains will report more accurate ETA records during the vessel's approach to the port as these factors become clearer and more stable.

3.4.3 Statistical analysis on vessel delays

Because a vessel may report several ETA records when it is sailing to the port, and each ETA record has a different degree of inaccuracy, we choose the last reported ETA record of each visiting vessel to analyze vessel delay by calculating the absolute difference between the vessel's ETA data and ATA data for one voyage. A negative value of vessel delay shows that the vessel arrives later than estimated, a positive value indicates that the vessel arrives earlier than estimated, and a value of 0 indicates that the vessel arrives on time. In this way, we can classify ship arrival data into the following three classes: early arrival, on time, and late arrival. There are 26 data records in the on-time class. For the other two classes, we analyze the minimum, maximum, median, and mean vessel delay in each class, and we calculate the standard deviation of the delay for all of the visiting ships. The vessel delay analysis results (in hours) are summarized in Table 4.

Table 4: Vessel delay analysis in hours

Type	Total records	Minimum	Maximum	Median	Mean	Standard deviation
Late	11,809	0.1	117.9	0.2	4.8	9.8
Early	1,571	0.1	34.1	1.1	2.8	4.5

In Table 4, there is a very large value in the Maximum column for late arrival ships: 117.9. The main reason for this unexpected value is that the ship's status was "pending" for a few days before it changed to "approved." This verification process delayed the ship's arrival to the port to much later than expected.

We take the absolute value of vessel delay to analyze the degree of the delay trend, and the vessel ETA record delay rate is defined as follows:

$$\text{Delay\_rate} = \left|\frac{A-E}{A-R}\right|, \tag{1}$$

where A represents the ATA, E represents the ETA, and R represents the ETA report time. The formula reflects the ratio of the actual delay error to the difference between the ATA and ETA report time. The lower the ratio, the more accurate the time of the ship's arrival. Specifically, the denominator in Eq. 1 is used as a normalizer: for the same difference between ATA and ETA, a report is more meaningful if the difference between the report time and the ATA is greater (i.e., the ship is further from the port) than if the difference is smaller (i.e., the ship is closer to the port). We also analyze the minimum, maximum, median, mean, and standard deviation of the delay rates of all of the ETA records, and the results are shown in Table 5.

Table 5: ETA delay rate analysis

Total records	Minimum	Maximum	Median	Mean	Standard deviation
23,547	0.0	1,349.0	0.19	0.38	8.8

In Table 5, there is also an extreme value in the Maximum column at 1,349.0 whose cause is similar to the cause of the extreme value in Table 4: severe vessel delay resulting from the trip's overly long time in "pending" status. We further assume that a vessel is punctual if its absolute delay (which refers to both early arrival and late arrival) is within 1 hour. In that case, only 16.3% (2,195) of the vessels visiting the HKP in 2021 arrived punctually. More analysis of the punctuality of ship arrival is conducted in the next section.

4 Evaluation of vessel arrival punctuality

4.1 Model assessment metrics

For the offline evaluation of ship arrival punctuality for one voyage, the ETA data are regarded as the predicted value and the ATA data are regarded as the ground truth value. To comprehensively assess the punctuality of a vessel's arrival to the HKP, five common metrics are adopted [25]: the RMSE, the mean squared error (MSE), the mean absolute deviation (MAD), bias, and the MAE. Given a total number of n ships, y_i is the ground truth ship arrival time (ATA), $\bar{y}$ is the mean value of the given dataset (ETA or ATA), and $\hat{y}_i$ is the predicted ship arrival time (ETA) for ship $i, i = 1, \ldots, n$, the definitions of the metrics are as follows:

RMSE:

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

MSE:

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}. \tag{3}$$

MAD:

$$\text{MAD} = \frac{\sum_{i=1}^{n} |y_i - \bar{y}|}{n}. \tag{4}$$

Bias:

$$\text{Bias} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n}. \tag{5}$$

MAE:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| . \tag{6}$$

These five metrics are used in the following sections to comprehensively assess ship arrival punctuality derived from public data and the accuracy of the prediction of ship arrival time to the HKP. The evaluation results can help port authorities analyze ship arrival status more efficiently and thus better allocate port resources for vessel services, improving the port's service level and competitiveness.

4.2 Data discretizing

The accuracy of ETA data reported by vessels on their way to the HKP cannot be directly compared if they are within different time boundaries. For example, the accuracy of an ETA record reported by a vessel 30 hours before arrival is highly likely to be much lower than the accuracy of an ETA record reported by a vessel 6 hours before arrival, as the influences of uncertain factors along the way (e.g., sea and weather conditions and navigation status) dissipate as the ship approaches the port, and more accurate ETA data are therefore expected to be reported. Accordingly, it is unfair to directly compare the accuracy of these two ETA records given the large difference in their report times. For this reason, it is necessary to discretize the ETA data into different time slices before comparing them. Because vessels arriving at the HKP start to report their ETA data to the port 36 hours before arrival, we first calculate the difference between the ETA report time and the ATA of each historical ETA record in hours. Second, the time difference is divided into 37 time slices from "0 hour" to '≥36 hour," with 1 hour as the interval. For example, the "0 hour" time slice includes ships whose difference between the ETA report time and its corresponding ATA is between 0 and 1 hour. The "≥36 hour" time slice contains ships with a time difference of no less than 36 hours (e.g., 37 or 38 hours). The number of ETA records in each time slice is shown in Figure 5.

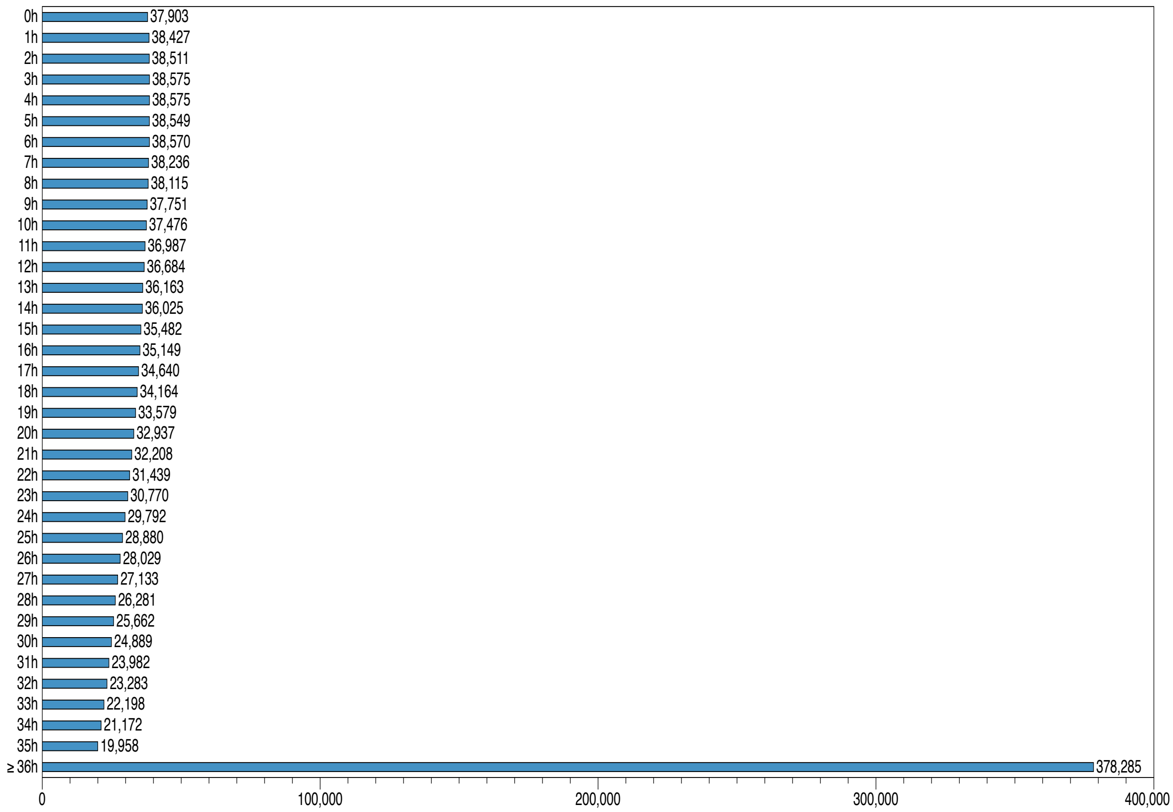


Figure 5: The number of ETA reports in each time slice

As shown in Figure 5, the "≥36h" slice contains the largest number of ETA reports (378,285), and the data volume drops dramatically to 19,958 records in the "35 hour" slice. There are two reasons for this phenomenon. The first is that vessels start to report ETA data at least 36 hours in advance in accordance with the requirements of the Hong Kong Marine Department. The second is that the "≥ 36 hour" slice includes ETA data where the difference between the ETA report time and the corresponding ATA is more than 36 hours. Consequently, this time slice is associated with the largest number of ETA reports. The number of ETA records then gradually increases from 19,958 records in the "35 hour" slice to 37,903 records in the "0 hour" slice as the arrival time approaches. Indeed, when vessels are approaching the port, its ETA becomes increasingly certain. Therefore, ship captains seek to provide the port with an updated ETA record so that the port can be better prepared for their arrival.

4.3 Evaluation result

The evaluation results of the punctuality of vessel arrival in different time slices are shown in Figure 6. The left coordinate of the figure is the values of the RMSE, Bias, MAE, and MAD in different time slices measured in units of 1 minute (min), whereas the right coordinate is the value of the MSE metric measured in units of min^2.

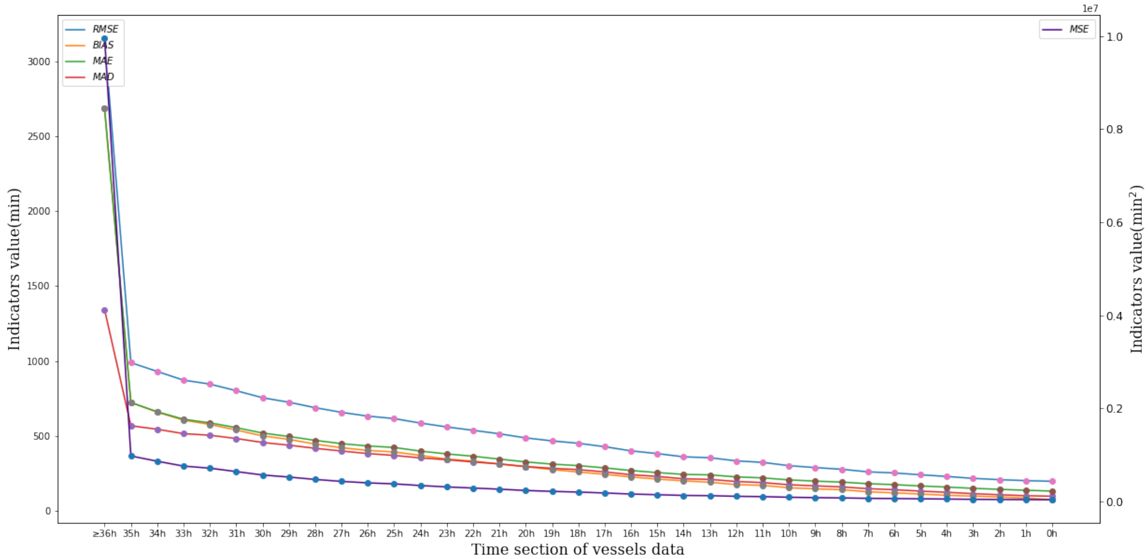


Figure 6: Punctuality of vessel arrival in different time slices

From Figure 6, we see that the overall error trend decreases as vessels approach the HKP, as shown by all of the evaluation metrics. In other words, the smaller the time difference between a vessel's report time and ATA, the more accurate the ETA data. This result shows that a vessel's ETA data become more accurate and reliable as the vessel approaches the port. Furthermore, it is interesting to witness a significant drop in the RMSE by 69% from 3,155 (min) at the "36 hour" slice to 989 (min) at the "35 hour" slice, and similar situations can be seen in the other four metrics, for which the error decreased sharply from the "36 hour" slice to the "35 hour" slice. The main reason is that the "36 hour" slice also includes ETA records where the time difference between the report time and the ATA is more than 36 hours. In earlier time stages, vessels tend to report ETA data that contain larger errors.

5 Prediction of vessel arrival time to HKP using data-driven models

This section aims to develop a data-driven model to predict ships' arrival time at the HKP. This section contains three main parts: feature engineering, vessel arrival time prediction, and analysis of the prediction results.

5.1 Feature engineering

Feature engineering is a series of engineering processes on the original data to extract and refine the features before inputting them into machine learning models. Here, we use four feature engineering steps (feature selection, data fusion, categorical feature encoding, and dataset split) to address time-related vessel data and vessel physical factors, which are clarified below.

5.1.1 Categorical feature encoding

Feature encoding on categorical features is an essential and vital part of the data engineering pipeline [39]. Categorical data are a kind of non-numeric data and are often divided into groups. For example, vessel type is a type of categorical data containing values such as container, bulk, LNG, and chemical tanker vessels and are stored in a string format. These features cannot be directly processed by machine learning algorithms. Therefore, categorical feature encoding is implemented to convert these strings to numbers that can be fed into machine learning models. In light of the nature of the categorical data involved in this study, we mainly use the following three feature encoding methods: one-hot encoding (OHE), which is used to encode unordered categorical features; label encoding, which is used to encode ordered categorical features; and binary encoding, which is used to encode binary features. The methods of applying the encoding methods are as follows:

a) OHE: For a feature with m categories without order between them, after OHE processing, that feature is extended to m new binary features and the original feature is deleted, with each new feature corresponding to a category. The m binary features are mutually exclusive, and only one of them is set to 1 considering the real feature value, with 0 given to all of the $(m-1)$ features [39].

b) Label encoding: In label encoding, we assign labels based on hierarchy. For a feature with m categories, after label encoding, each category is mapped to a number between 0 and $m-1$. The larger the assigned value, the higher the hierarchical category [39].

c) Binary encoding: In binary encoding, the categorical feature is first converted to an ordinal number, and then the numbers are transformed into binary code (0 or 1) [39].

5.1.2 Feature extension and data fusion

Vessel arrival data given by the Hong Kong Marine Department website contain only time-related information (e.g., ETA, ATA, and report times) for visiting ships. To obtain more ship specifications, two more external databases are used: the World Register of Ships (WRS) [30] and the MarineTraffic website [13]. The WRS database contains the features (e.g., IMO number, call sign, ship type) of more than 100,000 vessels. MarineTraffic is the world's leading provider of vessel factors, shipping tracking, and maritime intelligence [13]. Several vessel generic features (such as length, depth, beam, and GT) are selected from these two databases and combined with vessel arrival data by vessel IMO number.

5.1.3 Data preprocessing

In the prediction task, our model predicts the deviation of a ship's ATA from its given ETA, because ETA and ATA data are timestamps that cannot be directly forecast and thus prediction accuracy is difficult to quantify. The predicted deviations plus the ETA data give us the final predicted vessel arrival time. We first combine the vessel arrival data and the WRS and MarineTraffic databases into a uniform dataset. The combined dataset contains vessel time-related information and vessel generic features. After feature encoding and extension, there are 31 features in the dataset that are selected as model inputs, which can be divided into the following two categories: continuous features (4) and categorical features (27). The description of the continuous features and their statistical information for the full dataset are provided in Table 6, and the description of the categorical features is provided in Table 7.

Table 6: Description of continuous features

Feature name	Meaning	Min value	Max value	Average value
Beam (meter)	Width of the hull.	7.8	61.5	30.8
GT (100 cubic feet)	Measure of a vessel's overall internal volume.	118.0	228,786.0	38,766.9
Length (meter)	The overall maximum length of a vessel.	25.4	400.0	207.1
E-R (hour)	Absolute difference between ETA and its report time.	0.1	35.8	21.7

Table 7: Description of categorical features

Feature name	Meaning	Feature encoding
ETA_day	Week day of ETA. Monday (13.9%), Tuesday (13.9%), Wednesday (13.8%), Thursday (14.3%), Friday(14.8%), Saturday (15.1%), Sunday (14.2%).	One-hot encoding
Report_day	Week day of report time. Monday (13.9%), Tuesday (13.9%), Wednesday (13.8%), Thursday (14.3%), Friday (14.8%), Saturday (14.8%), Sunday (14.5%).	One-hot encoding
ETA_shift	Hour shift of ETA Shift 1 (36.1%), shift 2 (33.1%), shift 3 (30.8%). (Recall that shift 1 is 0:00 ∼ 8:00, shift 2 is 8:00 ∼ 16:00, and shift 3 is 16:00 ∼ 24:00)	Label encoding
Report_time_shift	Hour shift of report time Shift 1 (37.8%), shift 2 (31.4%), shift 3 (30.8%).	Label encoding
Vessel type	Vessel type in the ETA data Container (74.8%), bulk (7.4%), heavy lift cargeo (7.0%), tanker (3.0%), chemical or LNG (3.2%), others (4.6%).	One-hot encoding
Status	Vessel trip status Approved (88.1%), pending (11.9%)	Binary encoding

Finally, we randomly split the data into a training set (18,836 records) and a test set (4,710 records) with a proportion of 4:1. The feature engineering steps are summarized in Table 8.

Table 8: Summary of data preprocessing scheme

Step sequence	Data preprocessing method	Meaning
(1)	Feature selection	Selecting 31 features regarding vessel-related physical factors and vessel arrival historical data that are regarded as closely related to vessel arrival punctuality.
(2)	Data fusion	Combining vessel arrival time records from the Hong Kong Maritime Department website and vessel-related factors from the WRS and the MarineTraffic website using the ship's IMO number.
(3)	Categorical features encoding	OHE is applied to ship type, ETA, and report time, binary encoding is applied to trip status, and label encoding is applied to the hour shift of report time and ETA data.
(4)	Dataset split	Randomly splitting the full dataset into a training set (80%) and a testing set (20%).

5.2 Vessel arrival time prediction

We implement an RF model to predict the actual time of vessel arrival at the HKP. In the following sections, we will introduce RF model in details regarding model construction and evaluation.

5.2.1 Introduction of RF model

RF model is a type of ensemble learning algorithm based on CART and bootstrapping aggregation method [3]. A CART regression tree is constructed in a recursive manner where each node is split into two child nodes using MSE as the splitting criterion [10]. A simple decision tree with depth three (i.e., the number of layers from the root node to the deepest child node) is shown in Figure 7.

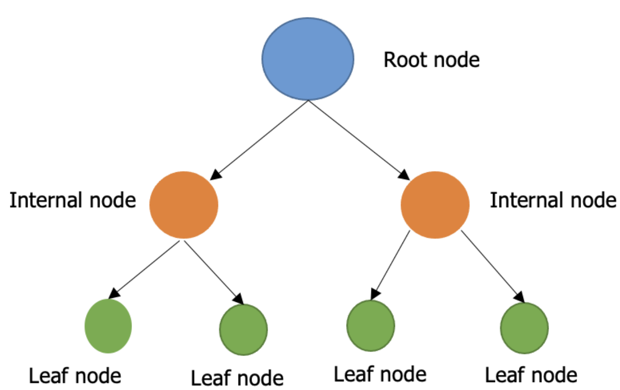


Figure 7: The structure of an example CART

The top node in a CART is called the root node. It includes all samples for training and it is split into subsequent nodes during the construction process in a recurrent manner. Node splitting for a regression task aims to make the samples in the subsequent nodes as similar as possible regarding their targets by minimizing the MSE of the subsequent nodes. For one split, a feature and one of its values are selected as the splitting point. When any preset tree growing termination condition is reached in this process, the corresponding node will not be split and it becomes a leaf node that gives the final prediction results.

To be more specific, suppose we have n samples in the training set:

$$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2) \ldots\ldots, (\boldsymbol{x}_n, y_n)\}, \tag{7}$$

where $\boldsymbol{x}$ is a d-dimensional feature vector and y is the prediction target. The tree splitting steps start from the root node. First, we select a feature d_i and one of its values s_i as a candidate splitting point denoted by (d_i, s_i) to split the whole dataset D into two sub-areas R_1 and R_2. These two areas can be represented by:

$$R_1 = \{y_i \mid \boldsymbol{x}_{i,d_i} \leq s_i\}, R_2 = \{y_i \mid \boldsymbol{x}_{i,d_i} > s_i\}. \tag{8}$$

The mean target values of all samples in R_1 and R_2 are set as the predicted targets of the samples contained in R_1 and R_2, respectively. Then, the sum of MSE of R_1 and R_2 is calculated as the performance of this candidate splitting point. All features and their values are then iterated to form all candidate splitting points to find the one that minimizes the sum of MSE values of the two areas, and the optimal split pair (d^*, s^*) leading to the minimum sum of MSE is selected as the final splitting point to divide the current node. Then, the above steps are repeated in each node until any tree growing termination condition is reached [24].

Specifically, mathematical representation of the above steps to construct CART regression tree is shown as follows:

1) Starting from the root node, select a feature value pair (d_i, s_i) to split the dataset D into two areas R_1 and R_2 which are represented by Eq. (8).

2) Calculate the mean targets of samples C_1 and C_2 of the two sub-areas by:

$$c_1 = \frac{1}{n_1} \sum_{\boldsymbol{x}_i \in R_1} y_i, c_2 = \frac{1}{n_2} \sum_{\boldsymbol{x}_i \in R_2} y_i, \tag{9}$$

   where n_1 and n_2 are the numbers of nodes in the corresponding areas.

3) Next, we iterate all features d_i and their values s_i to choose the split pair that can minimize Eq. (10). The optimal split pair of the current node is denoted by (d^*, s^*).

$$(d^*, s^*) = \min_{d_i, s_i} \left\{ \min_{\boldsymbol{x}_i \in R_1(d_i, s_i)} (y_i - c_1)^2 + \min_{\boldsymbol{x}_i \in R_2(d_i, s_i)} (y_i - c_2)^2 \right\}. \tag{10}$$

4) Divide the samples into two areas with the optimal splitting pair (d^*, s^*). The subsequent two new areas are:

$$R_1(d^*, s^*) = \{y_i \mid \boldsymbol{x}_{i,d_i} \leq s^*\}, R_2(d^*, s^*) = \{y_i \mid \boldsymbol{x}_{i,d_i} > s^*\}. \tag{11}$$

5) Execute steps 1 and 2 on nodes until reaching any of the present tree growing termination condition and no node can be further split. Nodes in the lowest layer become leaf node.

6) Finally, we divide the whole training set into M areas $R_1, R_2 \ldots\ldots R_M$ where M is also the number of leaf nodes. The generated decision tree can be represented by:

$$f(\boldsymbol{x}) = \sum_{i=1}^{M} c_i I(x \in R_i). \tag{12}$$

   Where $m = 1, ..., M.$ and I is a indicator function takes the following form:

$$I = \begin{cases} 1 & \text{if } (\boldsymbol{x} \in R_m), \\ 0 & \text{if } (\boldsymbol{x} \notin R_m). \end{cases} \tag{13}$$

However, traditional CART models suffer from the problem of overfitting, leading to weak generalization capability [3, 24], as they are sensitive to extreme data and subtle changes. To overcome this issue, a bootstrap aggregating (bagging) method is proposed to create divergence in the training set by using an ensemble of CART models to construct a unified model. The basic idea of CART with the bagging method is presented as follows:

1) Suppose that we have an original training set with n samples. To form a bootstrap sample, n samples from the original training set are randomly extracted with replacement. Then, this process is repeated k times [3], and we have a total of k bootstrap samples after the resampling process [3].

2) Train k CART models using the K bootstrap samples. In the regression problem, the final output is given by averaging the outputs of the k CART models.

The RF improves on the bagging method based on the CART model. The only difference between RF and CART models with bagging is the manner in which each node in the tree is split. The optimal split pair is selected from a random subset of features instead of all of the features in the RF model, and the number of selected features is preset. With this characteristic, the RF model can handle high-dimensional data without feature selection and is more robust against overfitting [3]. In maritime studies, the RF model is widely used for vessel fuel consumption prediction, ship energy efficiency prediction, and the efficient inspection of vessels, among other topics [33, 35, 38, 36, 38, 26, 34]. Here, we develop an RF regression model implemented by the scikit-learn machine learning library in Python [19, 19] to predict a vessel's ATA to the HKP based on vessel historical arrival data, vessel ETA data, and vessel physical characteristics.

5.2.2 Hyperparameter tuning

Hyperparameters have a large impact on machine learning models, and their values should be set before model training. The RF model has several hyperparameters, and we tune the values of max_depth, min_samples_leaf, min_samples_split, n_estimators, and max_features. According to the RF document in scikit-learn [19], the definitions and default values of the selected hyperparameters are listed in Table 9.

Table 9: Hyperparameters to be tuned in the RF regression model

Hyperparameter	Meaning	Typical default values
max_depth	Maximum depth of each CART in the RF model	None
min_samples_leaf	Minimum number of examples permitted to be contained in a leaf node	1
min_samples_split	Minimum number of examples in a node less than which the node cannot be further split	2
n_estimators	Number of trees in the RF model	100
max_features	Number of features considered for splitting a node in each tree of the RF model	'sqrt'

The "none" value for hyperparameter max_depth in Table 9 means that there is no limitation on the depth of the tree, i.e., the nodes can be expanded until there is only one sample in each leaf node or all of the leaves are pure (i.e., with samples that have the same output). We implement a grid search and the K-fold cross-validation (K-fold CV) method to find the best values for the hyperparameters shown in Table 9. A grid search is a tuning method that exhaustively searches combinations of the candidate values for all candidate hyperparameters to find the set of hyperparameter values that leads to the best performance on the validation set(s) [3, 2]. The search ranges and intervals of the hyperparameters for the RF model are listed in Table 10.

Table 10: Range of values for the specified hyperparameters

Hyperparameter	Range	Interval
max_depth	From 1 to 100 or None	6
min_samples_leaf	From 2 to 12	2
min_samples_split	From 3 to 13	2
n_estimators	From 1 to 1000	20
max_features	"Auto" or "MSE"	\

The "auto" in max_feature in Table 10 means that we consider all input features when finding the best splits, and "sqrt" means that we only consider the square root of the number of features for each node splitting.

In K-fold CV, we first split the training set into K subsets, where each subset is called a fold [9]. Then, one subset is used as the validation set and the remainder $(K-1)$ subsets are used as the training set. Each subset should be used as the validation set, and thus the above process is repeated K times. Finally, we average the performance of all folds when they serve as the validation set to generate the final validation result. If we fit the model with five-fold CV, an illustration of the hyperparameter tuning process is shown in Figure 8.

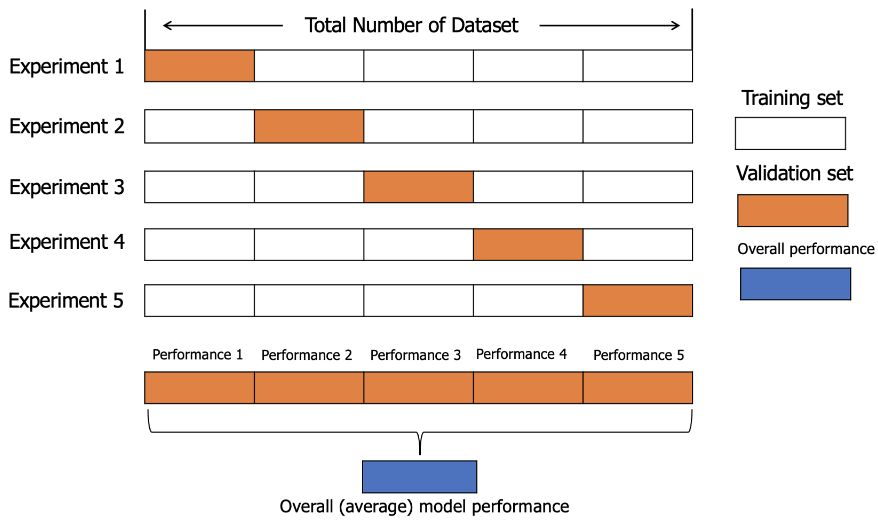


Figure 8: An illustration of five-fold CV

We tune the hyperparameters with a grid search and five-fold CV to obtain the optimal parameters. Specifically, we first generate a grid that contains the sets formed by all of the candidate values for all of the hyperparameters, and then we tune and choose the set of hyperparameters that has the best performance by iterating all possible sets in the generated grid with five-fold CV. The values that are ultimately adopted for the hyperparameters are listed in Table 11.

Table 11: Adopted hyperparameters for the RF regression model

Hyperparameter	max_depth	min_samples_leaf	min_samples_split	n_estimators	max_features
Selected value	“None”	10	7	661	“Auto”

5.3 Analysis of prediction results

The final RF model is trained on the full training set using the hyperparameter values given in Table 11. Because the model output is the predicted deviation value, the sum of the original ETA data and the deviation is the vessel’s predicted arrival time. We repeat the above training and evaluating steps 10 times to reduce random errors, and the min, max, mean, and variance of the RMSE/MAE values on the datasets are shown in Table 12. The first five rows, i.e., the “Predicted results” in Table 12, correspond to the performance of our RF model’s ship arrival time prediction. The last five rows, i.e., the “Test set,” correspond to the deviation of the ETA data in the testing set, which is the difference between the reported ATA and the reported ETA.

Table 12: A comparison of the proposed RF model and the original test set

Data	Metric	Minimum	Maximum	Mean	Variance
Predicted results	RMSE	15.251	15.295	15.272	0.00005
Predicted results	MAE	13.775	13.783	13.780	0.000007
Predicted results	Bias	13.696	13.726	13.701	0.007
Predicted results	MAD	5.481	5.483	5.481	0.0005
Predicted results	MSE	233.022	233.025	233.023	0.006
Test set	RMSE	25.543	25.543	25.543	0
Test set	MAE	13.786	13.786	13.786	0
Test set	Bias	13.899	13.899	13.899	0
Test set	MAD	15.390	15.390	15.390	0
Test set	MSE	652.223	652.223	652.223	0

The prediction results in Table 12 show that the RF model achieves good prediction performance on the testing set when using the RMSE, MAD, and MSE as evaluation metrics. The delay error in the RMSE decreases from 25.5h in the original testing set to 15.3h in our predicted model, which reduces vessel delay errors by 40%. With respect to the MSE evaluation metric, the delay error decreases 65% from 652h^2 to 233h^2, representing a drop of approximately 64%. With respect to the MAD, the error decreases from 15.4h to 5.5h, which is approximately 64%. Next, we test and analyze the performance of the pre-set model in different time slices. Based on the ETA data distribution in Figure 5, we classify our dataset into three sets. Set 1 contains the data from the '0 hour" slice to the "17 hour" slice in Figure 5, set 2 contains the data from the "18 hour" slice to the "35 hour" slice, and set 3 includes the data in the "$\geq$36 hour" time slice. The prediction results are shown in Table 13, where Prediction 1 indicates our model prediction performance in set 1 and Testing set 1 corresponds to the original set 1.

Table 13: Prediction results in different test sets

Data \ Metric	RMSE	MAE	Bias	MAD	MSE
Prediction 1	3.71	1.56	1.40	1.39	13.80
Test set 1	4.22	1.63	1.45	1.88	17.81
Prediction 2	7.00	2.11	3.91	3.61	49.13
Test set 2	7.821	2.21	4.01	4.62	60.96
Prediction 3	33.09	5.40	29.63	12.15	1094.76
Test set 3	40.53	5.41	28.64	22.92	1642.02

Compared with the prediction results on the original ETA dataset shown in Table 12, the prediction results on the subsets in Table 13 show a similar trend: the RF model performs well when evaluated using the RMSE, MAD, and MSE metrics: the delay error drops by 12% in Testing set 1 from 4.2h to 3.7h, by 10% in Testing set 2 from 7.8h to 7h, and by 18.5% in Testing set 3 from 40.5h to 33.1h when evaluated using the RMSE. With respect to the MAD, the error decreases by 26% from 1.88h to 1.39h in Testing set 1, by 22% from 4.6h to 3.6h in Testing set 2, and by 43.6% from 22.9h to 12.1h in Testing set 3. The results in Table 13 show that the trained RF model is the most effective for set 3: The further the ship is from the port, the better the prediction. In contrast, for the MAE and Bias, there is no significant difference between the proposed RF model and the original dataset in both tests, as shown in Table 12 and Table 13. The reason for this result is that the loss function in the RF model is the MSE, which is expected to minimize the MSE, RMSE, and MAD in the prediction results, whereas the MAE and Bias may not necessarily be notably changed during the training process. To explore the performance of the RF model evaluated using the MAE, we train the RF model with "absolute_error" as the loss function. The prediction results show that for all of the datasets, the error evaluated by MAE decreases by 20%, from 13.8h to 11.0h. In contrast, setting "absolute_error" as the loss function has a significant shortcoming: the training time will be very long. According to the documentation for the the Sklearn library [19], training the RF model with "absolute_error" is significantly slower than using "squared_error," because "absolute_error" is not continuously derivable when optimizing the model [3]. In our practical situation, training the RF model with "absolute_error" takes 100 times longer than training the RF model with "squared_error." After training and testing, we analyze the importance of features in the RF regression model developed. The RF feature importance score can be automatically calculated using a built-in function, rf.feature_importances, in the scikit-learn library [19], and the score reflects the importance of each feature to predicting the target. The higher the score, the more important (i.e., the higher the contribution of) the feature to the final prediction [19]. The top 10 important features and their scores are listed in Table 14.

Table 14: Top 10 numerical feature importance scores for the RF model

Variable	Importance score	Ranking
E-R(h)	0.215	1
Length	0.213	2
GT	0.201	3
Beam	0.131	4
Report_slice	0.050	5
ETA_slice	0.040	6
Trip status	0.029	7
ETA_Mon	0.019	8
ETA_Tue	0.016	9
ETA_Wed	0.011	10

Table 14 shows that the reported ETA minus its report time (E-R(h)), vessel physical characteristics (length, GT, beam), time slices of report time and ETA (Report_ slice, ETA_slice), and vessel trip status (Trip status) are crucial features to determining a vessel's arrival time at a port. Feature E-R(h), i.e., the reported ETA minus its report time, is the most important feature for ship arrival time prediction as an external factor, which can be explained by our findings in Section 3. The smaller the time difference between a vessel's report time and its ETA at the port, the more accurate the ETA record. Vessel generic features, an internal factor, also affects ships' arrival time at the port. Vessel physical features such as length, GT, and beam place second to fourth on the feature importance ranking. Vessel physical features directly affect port operational efficiency, berth occupancy, and terminal resource allocation. Therefore, these elements will eventually affect the prediction performance of the model. [17]. Moreover, the status of a vessel's trip to the HKP and the time slices of the ETA and report time are important to the prediction target. A vessel can only call the port when it is in an approved state. With respect to the time shift feature, terminal operating costs are lower and port operational efficiency is higher during normal working hours (Shift 2). Accordingly, the time slice feature reflects port operating conditions during a specified period.

5.4 Model extension and insights

The extensions and insights derived from the RF vessel arrival time prediction model proposed in Section 5 are summarized in Table 15 and are further explained below.

Table 15: Summary of extensions and insights from in prediction results

General perspective	Detailed implementation scenarios
Port operations insights	1) Quantitative evaluation of vessel delay 2) Efficient vessel arrival prediction 3) Port congestion reduction based on the prediction results
Rational commercial decisions	1) Vessel and port operators: a win-win approach for commercial decision-making 2) Port authorities: managing port resources more efficiently to increase benefits gained 3) Owners: reducing vessel operating costs and saving fuel
Advisable policy proposal	1) Require vessels to update their ETA within a specified time interval 2) Require vessels to upload more credible ETA data 3) Require vessels to upload their generic features alongside their ETA for more efficient planning

5.4.1 Port operations insights

First, this study gives insights into the punctuality of vessel arrival to a port in a quantitative manner, which helps operators assess vessel delay more accurately. Compared with the original ETA data uploaded by vessel operators, the proposed vessel delay evaluation model and vessel arrival time prediction model for the HKP are more solid and effective to assess and estimate vessel arrival time. Our evaluation model can help port operators evaluate vessel arrival punctuality at the HKP and more efficiently allocate the port's limited resources. Our prediction results suggest that the time interval between the ETA and its report time and vessel generic features are vital for the punctuality of vessel arrival at the HKP. The ship arrival time prediction model considering these features can reduce more than 40% of the delay error in terms of the RMSE. By taking the prediction results into account, port operators can schedule vessel arrival and port management in advance to reduce congestion. In addition, they can intelligently allocate berths and more efficiently manage quays and terminals for vessels.

5.4.2 Rational commercial decisions

Vessel arrival delay will increase both handling time and port operations costs [4]. Managing vessel arrival delays in a quantitative way creates a win-win situation for ship owners and port management authorities. From the perspective of vessel owners, less uncertainty surrounding vessel arrivals will reduce the waiting and operating time required for a vessel to call a port. Accurate ship arrival time predictions will prevent additional fuel losses and decrease ship operating costs, increasing profits for vessel owners [4, 38]. For port operators, trusted vessel arrival information will help them arrange vessel movements in advance and thus reduce port congestion [4]. In addition, they will increase the efficiency with which they manage port resources if they have reliable vessel arrival information and handle an increased number of ships in a given period, which will give the port a good reputation, better commercial value, and higher profits [4].

5.4.3 Advisable policy proposals

Feature "E-R(h)" is the most important feature of our ship arrival time prediction model using RF. Figure 6 in Section 3 shows that most of the vessels calling at the HKP make only one ETA data report during their sailing. To expand the dataset and improve our model prediction accuracy, the first policy implication of our work is that the Hong Kong Marine Department should make it compulsory for ships to report ETA data within a specified period, e.g., vessels must update their ETA data every hour prior to arrival. According to the data analysis in Section 3, most vessels (55%) do not update their ETA reports when approaching the port, and these data lack credibility when ships sail to port. With a mandatory updating policy, it can be expected that the vessel arrival dataset will be larger and more reliable. Second, based on our analysis of vessel arrival, it should be noted that some vessels report their ETA data without ever arriving at the HKP, which can create confusion. The Hong Kong Maritime Department could regulate its policy on vessels' ETA uploading to eliminate this type of error. For example, vessels such as tugs, which only pass through Hong Kong waters but do not visit the HKP, would not need to upload ETA reports. Furthermore, the RF model suggests that vessel generic features are essential for ship arrival time prediction. We collect the related features from a third-party dataset because they are not provided by the online data source. To simplify the process and enhance data credibility, we recommend that vessels upload vessel generic features alongside their ETA reports. In addition, the shipping industry has highlighted its deeper integration with artificial intelligence in recent years, and it is believed that this study can shed light on the promotion of intelligent shipping by the Hong Kong Maritime Department.

5.5 Further research

This study is the first attempt to evaluate and predict the punctuality of vessel arrival at the HKP. In the prediction section, time-dependent parameters and external vessel generic features are considered to develop our RF prediction model. However, weather conditions (e.g., air temperature, tidal information, wind speed, and port operating conditions) are not considered or evaluated in this study. First, for further research, the above factors could be considered and combined in our dataset to improve the model's prediction accuracy. Second, to improve prediction accuracy, several novel and state-of-the-art machine learning regression algorithms (e.g., XGBoost and LightGBM) could be implemented for the ship arrival time prediction task. Furthermore, smart prediction and then optimization methods could be proposed to derive more efficient port operating decisions [34, 32]. For example, we could first predict the key unknown parameters in a subsequent optimization model using machine learning algorithms while considering the structure and property of the optimization model at the prediction stage and then solve the optimization model to obtain the optimal decision. Specifically, in this study we predict a more precise vessel arrival time at the HKP. Therefore, the results could be used as key parameters to improve port operational efficiency.

6 Conclusion

In daily port operations, vessel arrival time uncertainty brings about disturbances, reducing the efficiency of port operations and causing economic losses. To evaluate and resolve these issues, we first apply a quantitative method to evaluate the punctuality of vessel arrival at the HKP. Our model shows that the overall delay decreases as vessels approach the HKP. Next, a data-driven approach based on an RF regression model is developed to predict vessel arrival time at the HKP. The prediction model can reduce the error in the predicted ship arrival time at the port by 40% (from 25.5h to 15.5h) when evaluated using the RMSE compared with the reported ETA data, and by 20% (from 13.8h to 11.0h) when evaluated using the MAE. The results of our predictions also show that the reported ETA minus the report time (i.e., E-R(h)), vessel physical characteristics (i.e., length, GT, beam), time slices of report time and ETA (i.e., Report_slice, ETA_slice), and vessel trip state (i.e., Trip status) are the crucial features that determine a ship's arrival time at port. The proposed vessel arrival time evaluation and prediction models are essential for port management and operations, and they provide a basis for future researchers to optimize the management of daily port operations.

This study sheds light on the advantages of quantitative assessment in the punctuality of vessel arrival and the precision of the RF model in vessel arrival prediction. Nevertheless, several research questions remain for further research in terms of evaluation, prediction, and optimization. In terms of evaluation, no studies have explored the impact of emergencies and natural disasters (typhoons, accidents at sea) on ship arrival time at port. One promising research topic involves quantitatively evaluating and estimating the influence of emergencies such as COVID-19 on the punctuality of vessel arrival. As a practical matter, we could compare indicators such as the number of vessel arrivals and the accuracy of ETA data at the HKP before and after COVID-19 during a single year. In terms of prediction, in addition to using the RF model for vessel arrival time prediction, other state-of-the-art tree-based methods, such as extreme gradient boosting (XGBoost), light gradient boosting machine (Light GBM), and Catboost, could be implemented to reduce training time and improve prediction accuracy. We could also combine more data sources to form more comprehensive datasets. For example, meteorological information including, but not limited to, temperature, tidal level, and wind speed, and port operating conditions, such as port congestion status and terminal/berth availability, could be collected and combined with historical vessel arrival data and vessel generic features. In terms of optimization, an optimization model for planning daily port operations, such as berth allocation and quay crane schedules, could be proposed to improve the efficiency of daily port operations. In addition, computational experiments could be conducted to numerically test the proposed optimization models.

References

[1] Alfredo Alessandrini, Fabio Mazzarella, and Michele Vespe. "Estimated time of arrival using historical vessel tracking data". In: *IEEE Transactions on Intelligent Transportation Systems* 20.1 (2018), pp. 7–15.

[2] Sylvain Arlot and Alain Celisse. "A survey of cross-validation procedures for model selection". In: *Statistics Surveys* 4 (2010), pp. 40–79.

[3] Leo Breiman. "Random forests". In: *Machine Learning* 45.1 (2001), pp. 5–32.

[4] Birgitt Brinkmann. "Operations systems of container terminals: a compendious overview". In: *Handbook of Terminal Planning*. Springer, 2011, pp. 25–39.

[5] Cheng-Chi Chung and Chao-Hung Chiang. "The critical factors: An evaluation of schedule reliability in liner shipping". In: *International Journal of Operations Research* 8.4 (2011), pp. 3–9.

[6] Hong Kong Government. *Vessel arrivals and departures*. 2022. URL: https://data.gov.hk/en-data/dataset/hk-md-mardep-vessel-arrivals-and-departures (visited on 05/23/2022).

[7] Sungil Kim, Heeyoung Kim, and Yongro Park. "Early detection of vessel delays using combined historical and real-time information". In: *Journal of the Operational Research Society* 68.2 (2017), pp. 182–191.

[8] Will Koehrsen. *Container service reliability stable in November*. 2015. URL: https://www.drewry.co.uk/news/news/container-service-reliability-stable-in-november/ (visited on 08/10/2022).

[9] Will Koehrsen. *Hyperparameter tuning the random forest in Python*. 2020. URL: https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74/ (visited on 07/22/2022).

[10] Roger J Lewis. "An introduction to classification and regression tree (CART) analysis". In: *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California*. Vol. 14. Citeseer. 2000.

[11] Chi-Chang Lin and Chia-Hsun Chang. "Evaluating employee's perception toward the promotion of safety marketing at ports". In: *International Journal of Shipping and Transport Logistics* 13.3/4 (2021), pp. 275–299.

[12] Meifeng Luo and Sung-Ho Shin. "Half-century research developments in maritime accidents: Future directions". In: *Accident Analysis & Prevention* 123 (2019), pp. 448–460.

[13] MarineTraffic. *Marine Traffic*. 2022. URL: https://www.marinetraffic.com/ (visited on 07/12/2022).

[14] Hong Kong Maritime and Port Board. *Port of Hong Kong Introduction*. 2022. URL: https://www.hkmpb.gov.hk/en/port.html.

[15] Takahiro Ogura, Teppei Inoue, and Naoshi Uchihira. "Prediction of arrival time of vessels considering future weather conditions". In: *Applied Sciences* 11.10 (2021), p. 4410.

[16] Claudia Pani et al. "A data mining approach to forecast late arrivals in a transhipment container terminal". In: *Transport* 29.2 (2014), pp. 175–184.

[17] Claudia Pani et al. "Prediction of late/early arrivals in container terminals-A qualitative approach". In: *European Journal of Transport and Infrastructure Research* (2015).

[18] Kikun Park, Sunghyun Sim, and Hyerim Bae. "Vessel estimated time of arrival prediction system based on a path-finding algorithm". In: *Maritime Transport Research* 2 (2021), p. 100012.

[19] Fabian Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *The Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[20] MD Robards et al. "Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)—a review". In: *Bulletin of Marine Science* 92.1 (2016), pp. 75–103.

[21] Nurul Haqimin Mohd Salleh et al. "Predicting a containership's arrival punctuality in liner operations by using a fuzzy rule-based bayesian network (FRBBN)". In: *The Asian Journal of Shipping and Logistics* 33.2 (2017), pp. 95–104.

[22] Shamika N Sirimanne et al. "Review of maritime transport 2019". In: *United Nations Conference on Trade and Development, Geneva, Switzerland*. 2019.

[23] Dejan Štepec et al. "Machine learning based system for vessel turnaround time prediction". In: *2020 21st IEEE International Conference on Mobile Data Management (MDM)*. IEEE. 2020, pp. 258–263.

[24] Vladimir Svetnik et al. "Random forest: A classification and regression tool for compound classification and QSAR modeling". In: *Journal of Chemical Information and Computer Sciences* 43.6 (2003), pp. 1947–1958.

[25] Albert Veenstra and Rogier Harmelink. "On the quality of ship arrival predictions". In: *Maritime Economics & Logistics* 23.4 (2021), pp. 655–673.

[26] NP Ventikos, P Sotiralis, and M Drakakis. "A dynamic model for the hull inspection of ships: The analysis and results". In: *Ocean Engineering* 151 (2018), pp. 355–365.

[27] Adrian Viellechner and Stefan Spinler. "Novel data analytics meets conventional container shipping: predicting delays by comparing various machine learning algorithms". In: *Proceedings of the 53rd Hawaii International Conference on System Sciences*. 2020, pp. 1278–1287.

[28] Shuaian Wang and Qiang Meng. "Sailing speed optimization for container ships in a liner shipping network". In: *Transportation Research Part E: Logistics and Transportation Review* 48.3 (2012), pp. 701–714.

[29] Richard T. Watson, Henrik Holm, and Mikael Lind. "Green steaming: A methodology for estimating carbon emissions avoided". In: *Proceedings of the International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015, Fort Worth, Texas, USA, December 13-16, 2015.* Ed. by Traci A. Carte, Armin Heinzl, and Cathy Urquhart. Association for Information Systems, 2015. URL: http://aisel.aisnet.org/icis2015/proceedings/Sustainability/1.

[30] WRS. *World Shipping Register.* 2020. URL: https://world-ships.com/ (visited on 07/10/2022).

[31] Xuhang Xu et al. "Trajectory clustering for SVR-based Time of Arrival estimation". In: *Ocean Engineering* 259 (2022), p. 111930.

[32] Ran Yan and Shuaian Wang. "Integrating prediction with optimization: Models and applications in transportation management". In: *Multimodal Transportation* 1.3 (2022), p. 100018.

[33] Ran Yan, Shuaian Wang, and Yuquan Du. "Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship". In: *Transportation Research Part E: Logistics and Transportation Review* 138 (2020), p. 101930.

[34] Ran Yan, Shuaian Wang, and Kjetil Fagerholt. "A semi-"smart predict then optimize"(semi-SPO) method for efficient ship inspection". In: *Transportation Research Part B: Methodological* 142 (2020), pp. 100–125.

[35] Ran Yan, Shuaian Wang, and Chuansheng Peng. "An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities". In: *Journal of Computational Science* 48 (2021), p. 101257.

[36] Ran Yan et al. "Analysis and prediction of ship energy efficiency based on the MRV system". In: *Maritime Policy & Management* (2021), pp. 1–23.

[37] Dong Yang et al. "How big data enriches maritime research–a critical review of Automatic Identification System (AIS) data applications". In: *Transport Reviews* 39.6 (2019), pp. 755–773.

[38] Jingjing Yu et al. "Ship arrival prediction and its value on daily container terminal operation". In: *Ocean Engineering* 157 (2018), pp. 73–86.

[39] Alice Zheng and Amanda Casari. *Feature engineering for machine learning: principles and techniques for data scientists.* " O'Reilly Media, Inc.", 2018.