

Article

Fuel Consumption Prediction and Optimization Model for Pure Car/Truck Transport Ships

Miao Su ¹, Zhenqing Su ², Shengli Cao ³, Keun-Sik Park ⁴ and Sung-Hoon Bae ^{5,*}

¹ The Graduate School of Technology Management, Kyunghee University-Global Campus, Yongin-si 17104, Republic of Korea; marksu@khu.ac.kr

² The Graduate School of Global Business, Kyonggi University, Suwon-si 16227, Republic of Korea; suzhenqing@kyonggi.ac.kr

³ School of Automation, Xi'an University of Posts and Telecommunications, Xi'an 710061, China; csl1993@xupt.edu.cn

⁴ Department of International Logistics, Chung-Ang University, Seoul 156-756, Republic of Korea; pksik0371@cau.ac.kr

⁵ Department of Trade and Logistics, International Logistics, Chung-Ang University, Seoul 156-756, Republic of Korea

* Correspondence: sunghoonbae@cau.ac.kr

Abstract: Predicting and optimizing ship fuel use is a crucial technology for lowering greenhouse gas emissions. Unfortunately, existing research is rarely capable of developing fuel consumption forecasts and optimization models for a particular transport system. This study develops a fuel consumption prediction model based on machine learning and a fuel consumption optimization model based on particle swarm optimization for ships. We studied nearly ten years of big data from a large Korean pure car and truck shipping company (PCTC), which contained 16,189 observations from 2012 to 2021. Results indicate that the XGBoost deep learning model outperforms conventional prediction models at the stage of fuel consumption prediction, with an R^2 of 0.97. Furthermore, in the fuel consumption optimization stage, the particle swarm optimization method can effectively reduce fuel consumption. This study helps PCTC companies control shipping costs and save energy. Insights for shipping businesses to meet environmental demands are provided as well.



Citation: Su, M.; Su, Z.; Cao, S.; Park, K.-S.; Bae, S.-H. Fuel Consumption Prediction and Optimization Model for Pure Car/Truck Transport Ships. *J. Mar. Sci. Eng.* **2023**, *11*, 1231. <https://doi.org/10.3390/jmse11061231>

Academic Editor: Theocharis D. Tsoutsos

Received: 1 May 2023

Revised: 26 May 2023

Accepted: 13 June 2023

Published: 15 June 2023



Copyright: © 2023 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: fuel consumption prediction; fuel consumption optimization; machine learning; maritime; XGBoost; particle swarm optimization

1. Introduction

The worldwide shipping sector is crucial to international trade since it transports 90% of all goods throughout the world [1,2]. However, the shipping industry faces two overwhelming problems due to adverse factors such as the global economic recession and climate warming [3]. First, the continuous increase in shipping fuel prices has increased operating costs for shipping companies. Second, during their voyages, ships emit large quantities of greenhouse gases (GHG), harming ecosystems, the weather, and human health in catastrophic ways [4]. Marine structures also exhibit a high degree of instability due to a wide range of environmental pressures [5]. According to the International Maritime Organization's (IMO) fourth GHG report (2020), the total GHG emissions from shipping increased from 977 million tons in 2012 to 1076 million tons in 2018 (a 9.6 percent increase). To increase revenue and enhance competitiveness, shipping companies must maximize the energy efficacy of their vessels and reduce their environmental impact [6].

However, increasing energy efficiency remains a contentious issue. It is challenging to reduce fuel consumption on existing ships by modifying the hull structure. Due to the relatively high cost of technological solutions, shipping companies have attempted to reduce their fuel consumption via a variety of operational means. Popular measures

include, for example, using weather routes and optimizing sailing pace to plan the voyage of the ship. For a fixed route during a voyage, one of the primary responsibilities of the shipping company is to plan the ship's daily speed in advance to ensure a timely arrival while minimizing the consumption of fuel. Optimizing a vessel's speed necessitates forecasting its fuel consumption under varying conditions. Ships' energy efficiency and their environmental impact may be improved by creating and using a ship fuel consumption prediction model, which has been the subject of much research in recent years [3].

Calculating a ship's fuel consumption under varying journey circumstances requires a set of flexible and complex models that capture the influence of many elements on the ship throughout the trip. In other words, an analysis of the overall trend observed in the data and the construction of a model to predict fuel consumption during a ship's voyage can assist decision-makers in identifying and correcting anomalies in the ship's fuel consumption. In addition, fuel consumption optimization models can guide shipowners in taking measures to improve energy efficiency by applying them to the route, speed, and leveling [7]. The ship's energy efficiency is maximized when crew or operators utilize fuel optimization tools in conjunction with fuel consumption forecast models [3].

Therefore, extensive analysis, modeling, and optimization of fuel consumption forecasting can significantly assist the management and operation of shipping companies and contribute to their sustainability [8]. Despite the growing corpus of research on the optimization and forecasting of ship fuel consumption, there are still enormous gaps [9]. First, there were inaccuracies in the vessel voyage data used to build the fuel consumption prediction model, as these datasets were derived primarily from manually completed vessel logbook data, such as noon reports. Optimizing fuel use requires an accurate, trustworthy, and real-time prediction of fuel usage. However, many modeling approaches are founded on a single model, but they lack precision and resiliency [10], and their effectiveness in practical engineering applications is unclear. Second, the majority of neural network-based machine learning models have been proposed for predicting ship fuel consumption, and their development often necessitates a substantial amount of training data. Their structure is mainly empirical, the parameters are difficult to tune, and the results of the prediction lack interpretability [9]. Third, traditional regression models for fuel consumption optimization require a large number of parameters and different modeling methods, are expensive to study, have low accuracy, and have less consideration of various parameters such as speed and port processing time [11]. Finally, the variables that influence ship consumption of fuel are complex. Although it is generally acknowledged that a ship's speed is the most significant factor influencing its fuel consumption, other factors can also have an effect [12–14]. This includes, but is not limited to, trim conditions, displacement and draft conditions, weather and sea conditions, as well as hull and propeller irregularities. The creation of customized machine learning models is an ideal and promising strategy for addressing these challenges. Models based on machine learning can process multidimensional input data and extract latent information from complex datasets. In addition, they frequently have a superior capacity to deal with chaotic data. Compared to conventional statistical regression models, machine learning models can handle higher-dimensional data (such as ship displacement conditions, sea and weather conditions, aspect, and sailing speed) and make more accurate predictions, providing a more solid foundation for the development of customized ship fuel consumption reduction models.

This study offers numerous contributions. First, we have devised an artificial intelligence-based energy efficiency decision system using machine learning. The system monitors and predicts the fuel consumption of a ship by modeling and analyzing the ship's voyage data (historical data provided by the shipping company is used in this paper; however, real-time data can also be collected via hull data collection equipment and uploaded to the cloud). Anomalies can be identified dynamically to aid in shipping management. Secondly, using actual shipping data, a two-stage prediction and optimization model for fuel consumption was developed. XGBoost was determined to be the optimal ship fuel consumption model in the initial phase. In the second stage, a particle swarm optimization technique is used to

determine the optimal speed, thereby minimizing the ship's total fuel consumption. Third, this research offers management and operational recommendations for increasing the energy efficiency of transportation companies. Finally, we demonstrate the superiority of the XGBoost approach for predicting ships' energy efficiency and the particle swarm method's power and applicability for optimizing ships' fuel consumption.

The investigation's remaining structure is as follows: Section 2 reviews previous work on modeling the energy efficiency of ships using machine learning and other techniques, and Section 3 details the approach that will be used for this project. Section 4 then describes the proposed model and the training and validation results. Section 5 is the conclusion and discussion. Finally, recommendations for further study are offered in Section 6.

2. Literature Review

2.1. Linear Regression-Based Black Box Model

A completely data-driven black box model (BBM) requires no prior knowledge of the ship's physics [15]. It has the capacity to learn from experience, which improves the precision with which fuel consumption forecast models for ships can handle chaotic data. This model has also led to the pervasive implementation of BBM in recent scientific research. On the basis of linear regression models, the traditional BBM statistical model is a widely utilized classical method for predicting the fuel consumption of ships. Several types of linear regression models (simple linear regression, multiple linear regression, segmented linear regression, etc.) are widely used in these types of investigations [16,17]. In order to anticipate a ship's fuel consumption in full or ballast situations, traditional BBMs often suggest multiple linear regression (MLR) models for ships using numerous input variables, i.e., operational and environmental historical data [18].

Further, [19] also examined eight sister voyage reports and hull maintenance data from the Aframax crude oil tanker fleet and conducted an MLR analysis of fuel consumption rates. [20] proposed a regression-based method for estimating container ships' fuel consumption rate. Additionally, [21] proposed a statistical analysis approach based on sensor data gathered on board the ship to automatically determine the ship's operating mode (port, maneuvering, or sailing) and also presented fuel consumption standards for modeling in the transportation mode. For the purpose of evaluating the fuel consumption of vehicle transport boats, [22] developed the only polynomial regression model in the PCTC research area (Please see Appendix A for other abbreviations in this paper). They checked the draft and displacement, the temperature, the wind speed and direction, and the roughness of the hull and the propeller.

However, despite their intuitiveness and interpretability, statistical models based on linear regression still have drawbacks. For example, parametric statistical models require assumptions about the data distribution before constructing the model, which may produce biases [9]. In addition, linear regression models tend to perform inadequately when complex and multicollinear data are involved; they are susceptible to noisy data [23,24]. The benefits of machine-learning-based BBMs have therefore received more attention from scholars.

2.2. Literature Review of Machine Learning-Based Black Box Models

Recent advancements in communication technologies, such as data collection and storage, have spawned an explosion of research on ship operation data to govern navigation performance [25]. Machine learning and deep learning are only two of the AI methods that underpin this study. Scholars increasingly use advanced algorithms to model ship fuel consumption [3]. Furthermore, ML can robustly learn despite noisy data [26]. Machine learning models can handle multidimensional input data and extract hidden information from complicated datasets to make more accurate predictions [27,28]. Research demonstrates that machine-learning models perform better than statistical models [29]. Similar to developing BBM based on statistical modeling, researchers subject machine learning models to data collection and pre-processing, followed by selecting and developing suitable ML models based on requirements and further pre-processing of the input data, if necessary.

In addition, the machine learning model requires hyperparameter optimization based on training and validation sets in order to enhance its generalizability [9].

Many studies have noted advances in the ML field concerning sophisticated learning algorithms and efficient pre-processing techniques. As a result, some studies developed machine learning models related to ship fuel consumption [25,30–32]. On the basis of the principles of data fitting, we can broadly categorize these investigations into four groups: Models derived from statistical learning (e.g., LASSO and RIDGE) [30]; instance-based models (e.g., SVR and KNN) [32,33]; models based on trees (e.g., RF and DT) [32]; and models based on neural networks (e.g., ANN and LSTM) [34]. Using machine learning models for fuel consumption prediction simplifies the influence of complicated situations on fuel consumption, which is one of the biggest advantages of doing so [9]. In addition, sophisticated machine learning methods can more accurately predict a ship's energy consumption, particularly for container ships [34].

Notable is the artificial neural network-based model for estimating ship energy consumption, which employs ship operating and operational data [28]. The development of an artificial neural network (ANN) model necessitates domain expertise and access to frequently unavailable technical information [35]. Although there has been some research using ML methods to predict ship fuel consumption, ML is currently still in the development and iterative phases. However, analysis using more advanced techniques, such as deep learning, is still limited for predicting PTCT ship fuel consumption. Researchers such as [3,36] have also suggested building deep learning-based prediction models when the quality and quantity of data allow. Therefore, to fill the research gap, we incorporated relevant deep learning algorithms into the PCTC fuel consumption prediction analysis in this work. In summary, a complete prediction model based on various ML methods can help researchers optimize shipping operations and accurately reflect ship emissions to determine the best ML method, thus improving the model's accuracy and, ultimately, its use for route optimization.

2.3. Literature Review of Ship Fuel Consumption Optimization

The most important step in attaining strategic objectives such as fuel savings and emission reduction is optimizing ship fuel consumption forecasts, not merely predicting them accurately [3]. Existing ships can save energy and reduce emissions by optimizing their operations (route, speed, leveling, etc.) without altering their structure [37]. As sailing speed is the most significant influence, small speed changes can significantly improve a ship's energy efficiency, productivity, and revenue [38]. As a result, several studies have concentrated on optimizing speed throughout the voyage to reduce fuel consumption [17,29,39–42]. In the shipping literature, the speed-power curve is the most common method for estimating fuel consumption [43]. These studies all employ speed as the key variable and investigate the environmental impact of ship fuel usage [11].

In addition, shipping companies can save money on fuel by selecting the most efficient routes for their ships. The field of shipping route selection has also seen a rise in study in recent years [44]. For example, Ref. [39] applied the shortest route problem to the case of discrete arrival times. Many computational results showed the superiority of the shortest path strategy for shipping route selection and fuel cost calculation. Ref. [45] used a discrete-choice model to examine the impact of emission control areas (ECAs) on the worldwide shipping industry (DCM). In addition, Ref. [46] focused on the issue of choosing the service frequency for long-distance leased lines. The researchers tackled the network issue as a mixed-integer nonlinear program using a branch-and-bound approach that produces an approximate solution close to the optimum answer after a limited number of iterations. To find the most efficient path for container ships to take while using the least amount of fuel, Ref. [44] suggested an algorithmic solution to the asymmetric traveler's problem (ATSP) based on a deep machine learning method. Inputs to the model include five variables: average wind speed, sailing time, vessel capacity, wind speed, and wind direction. The model's mean absolute percentage error (MAPE) was 5.89%, suggesting that the prediction results were about 95% accurate. Despite

this, Ref. [47] conducted an analysis of previous research on ship routing and scheduling challenges and found that there is a paucity of liner shipping research.

Some experts and academics have also analyzed other uncertainties affecting the movement of ships. In addition to sailing pace, a ship's displacement, gross tonnage, cargo condition, and ballast water can dynamically influence its fuel consumption. It has also been demonstrated that sea conditions, such as currents [48], waves, and swells [49,50], influence ship fuel consumption. Moreover, meteorological conditions have been identified as a significant factor influencing the unpredictability of ship movements. For instance, Ref. [51,52] investigated the impact of meteorological conditions on ship performance and presented a collection of regression models. Recent studies have shown that models developed by [20,29] that combine marine and weather data, such as wind direction and wind force, wave direction and height, and seawater temperature, are highly accurate in predicting ship fuel consumption. Combining ship navigation-related characteristics with sea state and meteorological conditions has tremendous predictive and management potential for ship fuel consumption.

When coping with combined sea conditions, directionality must also be considered, specifically the relative direction of wind and swell [5]. According to numerous IMO and DNV reports, longitudinal inclination optimization can save four to six percent (or even fifteen percent) of fuel consumption [6]. Furthermore, modeling the ship's resistance under various longitudinal tilting conditions to determine the minimum resistance can minimize fuel consumption [3]. As a result, part of the study employs WBM for ship-side optimization through sophisticated CFD methodologies [53]. In addition, Ref. [54] utilized a simpler empirical model (Holtrop–Mannen) to estimate the prospective benefits of longitudinal tilt optimization. Further, Ref. [29,35] demonstrate that optimized ship leveling based on a data-driven strategy can result in fuel savings.

However, Ref. [3] observes that, due to the difficulty of incorporating marine environmental factors, the validity of the optimization results is questionable. Second, the researchers note that the longitudinal inclination optimization model implies that the longitudinal inclination of a ship can change in real time. It is challenging for a watercraft to alter its longitudinal inclination frequently (e.g., every 5 min). Thankfully, scientists are collecting a ton of information on ship fuel use, which makes it easier to develop normalized optimization models using data [3]. Furthermore, optimizing fuel consumption is fundamentally a multi-objective optimization problem. In practical applications, difficulties involving optimization of several objectives are common, and the actual issues are often complex due to the past's lack of effective solution methods.

Although there is a growing corpus of research on predicting and reducing ship fuel consumption, there is a significant gap in the current literature. First, current models for estimating ship fuel consumption primarily target tankers, container ships, ferries, tugboats, and passenger ships. To our knowledge, however, no models integrating machine learning techniques have been proposed for predicting fuel consumption and optimizing speed for pure car or truck vessels [9,12]. Due to the varied properties and structures of ships, researchers cannot universally apply one fuel consumption prediction model [13]. Therefore, research must create bespoke prediction models for each vessel type to improve prediction performance [14]. Second, the ship's sailing speed is the most essential input variable used to predict fuel consumption. However, the data used in the majority of studies is derived exclusively from noon reports. The absence of additional influencing factors, such as weather and sea conditions, prevented the collection of more comprehensive and accurate data. Thirdly, the majority of machine learning models proposed for predicting ship fuel consumption are based on artificial neural networks. However, the development of artificial neural network models typically necessitates a large number of training samples, and their structure is primarily based on previous experience. In addition, it is challenging to adjust the parameters of artificial neural networks, and the results of their predictions are not interpretable. In addition, it is challenging to calculate the influence of individual input variables on output variables. Fourthly, pioneering research integrating ship fuel prediction models with optimization models to reduce fuel consumption and CO₂ emissions is scarce. Fifthly, existing ship fuel consumption

prediction models are weak at analyzing the uncertainty of ship motion. Deep reinforcement learning (DRL) provides research ideas for solving the modeling of ship motion recognition in complex scenarios.

To cover the void, this paper develops a two-stage model by analyzing 16,189 pieces of data from significant car and trucking operations in Republic of Korea from 2012 to 2021 and developing a two-stage model based on the independent variables (characteristics) ship type, distance, fuel price, speed, and port time. In the first phase, we used 15 different machine learning methods to forecast a ship’s fuel usage and discovered that the XGBoost deep learning model performed the best. It displayed high accuracy, robustness, and sound engineering application value. Using the XGBoost model as the base model, in the second stage, a particle swarm optimization model is used to determine the optimal speed and in-port cargo handling time for the vessel’s current sailing conditions and, based on the predicted results and actual conditions, to develop the optimal sailing plan and operating strategy to reduce fuel consumption and improve energy efficiency. The benefits of the two-stage models presented in this paper are their capacity to manage high-dimensional data and their ability to make more accurate predictions than conventional statistical regression models. It is more efficient and produces more interpretable results than other machine learning models, such as artificial neural networks, and the degree of influence of features on the target variable can be generated for feature selection. In addition, it is able to account for a greater number of irregularities and other factors, thereby improving the accuracy of ship motion and attitude analysis. It can also be applied to additional ship categories, ship owners, and routes. It is also essential for real-time ship safety assessments.

3. Research Methodology

3.1. General Framework

This study’s primary objective is to design a two-step strategy based on a data-driven approach, including recommended methods and specific details for reducing fuel consumption and ship emissions. Figure 1 illustrates the research framework. A significant Korean pure car and truck (PCTC) transportation company provided the big data used in this study, which has 16,189 observations covering the time period from 2012 to 2021. The obtained characteristic data were standardized before being saved in a database. Using the collected data, we constructed an XGBoost deep learning integrated petroleum consumption prediction model in the first phase of the two-step strategy. In the second stage, a particle swarm model for optimizing fuel consumption (PSO) was constructed. We optimized the ship’s speed based on the established model in order to reduce fuel consumption and emissions. The suggested method is broken down into three distinct phases, as shown in Figure 1: (1) data collection and pre-processing; (2) prediction model analysis; and (3) optimization model analysis.

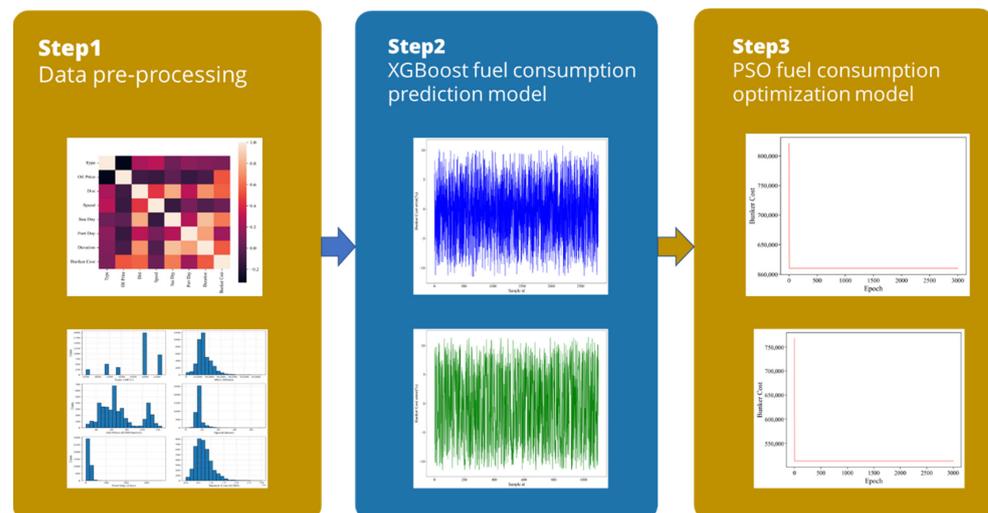


Figure 1. Research framework.

3.2. Modeling Methods

3.2.1. XGBoost Algorithm Framework

Extreme gradient boosting (XGBoost) is a sparse-aware technique for sparse data [55] that was first introduced in 2016 as a component of a scalable, robust tree augmentation machine learning (ML) framework. XGBoost’s basis is gradient-boosting decision trees (GBDT), which combine individual learners to generate dependencies through boosting. The classification and regression discipline makes extensive use of the XGBoost algorithm due to its quick, accurate, and efficient operations and robust generalization capability [56]. The main concept is to create a sample score by integrating the scores of each tree to produce a final prediction score for the sample, and then to learn new features by doing so. For example, the formula for predicting the score using K additive functions for n identifiers and m features is as follows:

$$\hat{y} = \sum_{k=1}^K f_k(x_i), f_k \in F \tag{1}$$

$$F = \left\{ f(x) = w_{q(x)} \left(q : R^m \rightarrow T, w \in R^T \right) \right\} \tag{2}$$

where F is the space of the regression tree, $f(x)$ is one of the regression trees, and $w_{q(x)}$ denotes each T , the independent structural score of the leaf tree. The following is an explanation of what is meant by the term “objective function” when referring to XGBoost:

$$L = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \tag{3}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \| w \|^2 \tag{4}$$

where l represents the loss function of the model, Ω is the regularization term, T denotes the number of leaf nodes, w is the fraction of leaf nodes, and γ and λ represent the control coefficients to prevent over-fitting.

When we generate the n th tree, we can write the predicted fraction formula as follows:

$$\hat{y}_i^{(t)} = \hat{y}_i^{t-1} + f_t(x_i) \tag{5}$$

where \hat{y}_i^{t-1} is the previous $t - 1$ round model prediction scores.

We can write the corresponding objective function as follows:

$$L(t) = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{6}$$

We employ Taylor’s second-order expansion to speed up the optimization process:

$$L(t) = \sum_{i=1}^n \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{i=1}^T w_j^2 \tag{7}$$

Then, the samples are recombined by adding the loss function of the samples, and finally, using the vertex formula to find the optimal w and the objective function formula L , we use the following equations:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{8}$$

$$L = -\frac{1}{2} \sum_{i=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{9}$$

$$G_i = \sum_{i \in I_j} g_i \tag{10}$$

$$H_i = \sum_{i \in I_j} h_i . \tag{11}$$

To find the best partition, XGBoost combines classical greedy and approximation algorithms, first listing a number of possibilities based on the percentile approach and then determining the best partition using Equations (8) and (9). Overfitting can be avoided using XGBoost thanks to its use of regularization, row sampling, and feature sampling, among other methods. It also has the ability to deal with sparse data. Parallel processing, one of XGBoost’s extra advantages, leads to a significant efficiency boost. In addition to its flexibility, the method has built-in cross-validation that permits cross-validation in every boosting iteration, as well as user-defined optimization targets and assessment criteria.

Scholars have applied it to disease prediction [57,58]; gene expression [59]; terrorist attack casualties [60]; industrial prediction [61]; and building engineering [62,63]. Yet, XGBoost is still seldom used for predicting ships’ fuel usage. We chose XGBoost as the model for forecasting fuel usage in this study by combining the aforementioned advantages of XGBoost with classification algorithms.

3.2.2. Particle Swarm Optimization Algorithm Framework

Kennedy and Eberhart first proposed particle swarm optimization (PSO) in 1995; avian predation behavior was the source of its inspiration. Each particle represents one possible solution to the goal function. The velocities of particles, which depend on both the particle’s and the population’s historical optimum solutions [60], determine where they are.

Assuming that the particle population contains n , the dimension of the search region is D dimensional. In addition, $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ is the particle i in D , the position of the particle in the dimensional space. $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ is the particle i , the velocity of the particle P_{best} is the individual extremum of the particle, i.e., the particle i found in the process of finding the optimal solution and the particle’s position in D . The particle’s coordinates in three space-time dimensions may be written as $P_i = (P_{i1}, P_{i2}, \dots, P_{iD})$. g_{best} is the optimal solution discovered historically by the entire population during the search process, whose position is in D . The position of the particle in the dimensional search space is $P_g = (P_{g1}, P_{g2}, \dots, P_{gD})$ for the first $k + 1$.

For the second iteration, the velocity and position of each particle in each dimension are iteratively updated based on the following formula:

$$\begin{aligned} v_{id}^{k+1} &= \omega v_{id}^k + c_1 s_{rand_1} (P_{id} - x_{id}^k) + \\ & c_2 s_{rand_2} (P_{gd} - x_{id}^k) \\ x_{id}^{k+1} &= x_{id}^k + v_{id}^{k+1} \end{aligned} \tag{12}$$

where v_{id}^{k+1} is the particle i in the first $k + 1$ generation, the d dimensional component of the particle d is the range of values of $[1, D]$; x_{id}^{k+1} is the velocity of the particle i in the $k + 1$ generation, which is the d dimensional component of the particle. Further, P_{id} is the value of particle i on the d dimensional component of the individual optimal solution; P_{gd} is the optimal solution for the whole population on the d -dimensional component. In addition, c_1, c_2 are the learning factors responsible for regulating P_{id} and P_{gd} , the maximum step size of the directional flight; $s_{rand_1} \setminus s_{rand_2}$ is the random number taken from $[0, 1]$ in Equation (1). v_{id}^k denotes the particle velocity value of the previous generation, $c_1 s_{rand_1} (P_{id} - x_{id}^k)$.

Individually optimal values are the consequence of the particle’s learning, which enables the particle to conduct a more effective global search and prevent falling into local optima; $c_2 s_{rand_2} \times (P_{gd} - x_{id}^k)$. The population learning component represents the capacity of elements within a population to share information with one another and the outcomes of population learning. Under the combined influence of these three factors, the whole population of particles iterates continuously, enhancing the development of the search area in a superior direction so that particles can seek the optimal global position.

In the particle swarm algorithm model, the model can consider an individual as a particle; then, the whole population is a particle swarm. Suppose, for instance, that an

n-dimensional target search space contains m particles, where we can write the *i*th particle's (*i* = 1, 2, ..., m) position as follows:

$$X_i = (x_i^1, x_i^2, \dots, x_i^n), (i = 1, 2, \dots, m)$$

Thus, the model can consider each particle position as a potential solution. By incorporating it into the objective optimization function, we can determine the position's or solution's optimality based on its corresponding fitness. If the particle is at its most advantageous location, we obtain:

$$P_i = (p_i^1, p_i^2, \dots, p_i^n), (i = 1, 2, \dots, m)$$

The best possible location for every particle in the whole particle population is the following: $P_g = (p_g^1, p_g^2, \dots, p_g^n), (i = 1, 2, \dots, m)$.

The particle's speed then becomes:

$$V_i = (v_i^1, v_i^2, \dots, v_i^n), (i = 1, 2, \dots, m)$$

In addition, the particle swarm algorithm uses the following formula to keep the positions of the particles updated:

$$\begin{aligned} v_i^d &= \omega v_i^d + c_1 r_1 (p_i^d - x_i^d) + c_2 r_2 (p_g^d - x_i^d) \\ x_i^d &= x_i^d + v_i^d \end{aligned} \tag{13}$$

ω is a positive number known as the inertia factor, c_1 and c_2 are non-negative constants called the acceleration constants or learning factors, and r_1 and r_2 are random numbers in the range [0, 1].

The formula for the speed increase has three parts on the right side of the equal sign: (1) the particle's "momentum" or "inertia" describes its propensity to continue moving at its current speed; (2) the "cognitive" indicates the particle's natural drive to optimize its past performance and reflects the particle's accumulated history; and (3) the "social" component, which indicates the particle's inclination to approach the group's or neighborhood's historical optimum, is informed by the group's historical experience of collaboration and information exchange among particles. A lower acceleration constant value permits particles to converge to their optimal solution more slowly, enabling a deeper exploration of the space of possible solutions between the present state and the best possible one. However, a too-low acceleration constant value may cause the particles to repeatedly fluctuate outside the optimal neighborhood and fail to search the target region effectively, resulting in reduced algorithm performance. A high acceleration constant value may lead to the particles repeatedly fluctuating outside the optimal neighborhood and failing to search the target region effectively. In most cases, $c_1 = c_2 = 2$ is used to denote the acceleration constant.

The amount of the initial velocity that is still being used is then represented by the inertia factor. If the inertia factor is significant, global convergence is stronger and local convergence is weaker. In contrast, if the inertia factor is smaller, local convergence is stronger and global convergence is weaker. Experiments demonstrate that the PSO algorithm converges quicker when using $\omega \in [0.8, 1.2]$ is used, so we chose $\omega = 1$ in this study. We restricted the range of position variation and velocity variation of the d-dimensional particle elements to $[x_{min}^d, x_{max}^d]$ and $[v_{min}^d, v_{max}^d]$, respectively. During the iterative process, if the position or velocity of a particle element in one dimension exceeds the set value, it is equal to the boundary value.

At the first step of the particle swarm method, all particles are given random beginning positions and initial velocities. Then, particles move forward in the problem space based on their velocities, their individual ideal positions, and the global optimal position. As the computation progresses, the particles aggregate or coalesce around one or more optimal points by exploring and exploiting favorable positions within the search space.

The technique is cleverly designed so that it remembers both the global optimal position and the particle ideal locations that have already been determined. In particular, we can summarize the PSO algorithm’s operation as follows:

- (1) The size, starting location, and beginning velocity of each particle are all part of the initialization process for a swarm of particles.
- (2) Find each particle’s fitness value using the objective function, then set the local and global optimum values to start with. Regarding the fitness function’s design, we may generate problem-specific designs. The core idea is that the size of the fitness value can determine whether the particle’s position is optimal.
- (3) Determine the termination condition’s achievement. If the goal is reached, the search process ends with the returned results. If not, proceed with the procedures that follow.
- (4) Change the velocities and positions of the particles in accordance with the formula for changing velocities and positions.
- (5) Determine the fitness of each particle according to the goal.
- (6) Refresh the global and local best values for each particle.
- (7) Set the termination condition of the iteration based on the specific problem, typically reaching the specified maximum number of iterations or the current optimal position of the particle swarm in order to satisfy the search requirements.

3.3. Ship Description

PCTC refers to ships that can transport a variety of ro-ro cargoes, including automobiles, buses, equipment, etc. PCTC is an enhanced PCC (pure car carrier) ship. The ship’s size is based on a Hyundai Accent Equivalent Unit (AEU) measuring 4.115 m by 1.62 m, or 6.66 square meters. The ship size in this study is between 3500 and 6700 AEU, and the shipload is between 13,000 and 19,000 DWT. There are between four and five liftable decks on a ship’s deck. One can raise and lower a liftable deck and alter its height. In addition, one uses a ramp to load and unload trucks, which we can conceptualize as a passageway between ships and land. The ramp capacity of the ships featured in this article ranges from 50 to 200 tons. There were various distributions of the ships from 1995 to 2017. Table 1 describes the key features of the case study vessels.

Table 1. Key features of case study vessels.

Ship Feature	Value	Unit
Vessel Size	3500~6700	[AEU]
Dead Weight	13,000~19,000	[Ton]
Liftable Deck	4~5	[Unit]
Ramp Capa	50~200	[Ton]
Built	1995~2017	[Year]

3.4. Description of the Data Set

The data used in this study is based on actual data from Shipping Company A (Republic of Korea) and equipment from Hyundai Group, Seoul, Republic of Korea. No company-specific information is provided in this study due to a confidentiality agreement with the shipping company. Data on various ship variables was sourced from different voyages.

A significant proportion of worldwide PCC/PCTC shipping comprises automobiles of South Korean origin [12]. Consequently, we gathered a dataset from a large PCC/PCTC shipping company in Republic of Korea, comprising ship statistics from 2012 to 2021. However, the overall flight data permitted only direct operations and time charter ships with records of shipping corporations operating ships because the actual shipping business did not implement or could not separate the two.

This study identifies approximately 37 ship routes based on the itineraries of PCTC vessels, which represent the majority of routes worldwide. However, considering that there are too many factors, such as the freight rate level, different shipping points, and

optimization modeling of each route, we picked three routes for optimization modeling in the second stage of fuel optimization. These three routes are, in order, Asia North West America (ASNW), Asia Arabian Gulf (ASAG), and Asia Europe (ASEU). We chose these three because, based on the raw data, the PCTC ships mostly use these routes. In addition, the usual routes for PCTC ships on these routes include the United States, Europe, and the Middle East. The choice of variables in machine learning is important because it affects the performance of the model and the speed of computation, leading to a better understanding of the process of data generation. Therefore, for fuel cost forecasting, we used independent variables such as vessel size, distance, speed, port day, and oil price (see Table 2). However, the scale of the ship does not alter over time. Therefore, it is not utilized as a variable in the time series.

Table 2. Variables for the research.

Variables	Abbreviation	Unit
Oil Price	Oil Price	USD
Distance	Dist	Miles
Speed	Speed	knot
Sea Day	Sea Day	Day
Port Day	Port Day	Day
Duration = (Sea + Port day)	Duration	Day
Bunker Cost	Bunker	USD

The speed of a ship is measured in knots, while loading and unloading distances are measured in miles. 380CST HSFO is used for voyages, and MGO is used in ports. The price of ship fuel is related to Brent crude oil prices, but it is not directly used to calculate the cost of ship fuel. Instead, the cost is determined based on a barrel of Brent crude oil, which is widely recognized in the global crude oil market. International oil prices are used to calculate ship fuel prices, taking into account differences in fuel supply and prices at different ports.

We calculated the fuel cost by multiplying the final cost by the quantity and price (price per ton) of fuel oil used during the voyage and diesel used during the breakdown. The ship’s continuous sails, the fuel oil from the previous voyage, any remaining fuel after the trip, and the unit price of each kind (fuel oil or diesel oil) make it difficult to distinguish between journeys. Therefore, we used the total fuel cost for analysis in this study. “Bunker cost” is the target value that we want to predict.

Researchers often use correlation as a preliminary technique for finding relationships between variables in machine learning, and it may be the key to improving the accuracy of predictive models. A correlation heat map is a graphical depiction of the correlation between numeric variables and illustrates the connection that exists between a number of different variables. The values in each cell represent the nature of the link between the two entities, with higher values suggesting a stronger bond and lower values indicating a weaker one. We can use positive or negative correlations between feature values to discover how independent features affect intuitive predictions. In most cases, a high positive correlation is shown when the Pearson correlation coefficient is larger than 0.7 [57]. Figure 2 shows the correlations between the dataset’s characteristics.

According to Figure 2, the results show that the correlation between some factors is very low; for example, the correlation between type and oil price is below 0.2. The reason for this is that the correlation between the price of oil (Brent) and the size of the ship (type) is very low in terms of intrinsic properties. However, when we analyze the cost of fuel used in the actual voyage of the ship, we can observe a correlation with the number of days sailed, duration, etc.

The low correlation between oil prices and ship type, distance, and speed is due to the varying fuel efficiency and cargo capacity of different ships, as well as the impact of weather conditions and navigational challenges on fuel consumption. The relationship between sea days and speed and port days and oil prices is also relatively low due to the

influence of other factors such as safety concerns and port charges. Overall, the impact of oil prices on shipping costs is complex and multifaceted, with many factors beyond just fuel costs influencing the final cost of transport.

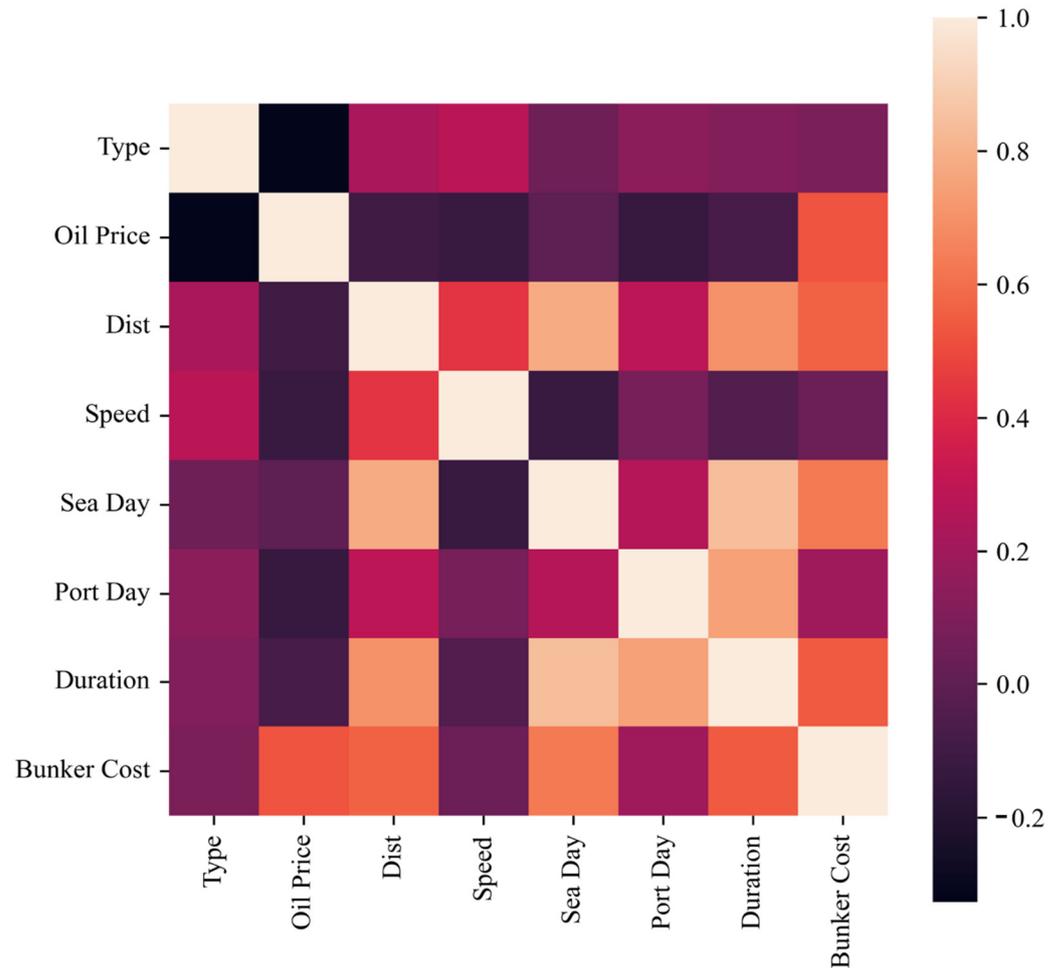


Figure 2. Correlation between data.

After detecting eigenvalue correlations and receiving help from an expert, we eliminated some eigenvalues with relatively high correlations. The remaining data includes ship type, distance, fuel price, speed, bunker cost, and port date. Table 3 describes the dataset’s features.

Table 3. Features in the dataset.

Variable	Unit	Feature Characteristic	Value of Feature
Vessel type	AEU	Integer	3500–6700
Speed	knot	Float	0.59–93.31
Sea Day	Day	Float	1.65–123.58
Duration	Day	Float	4.33–409.76
Oil price	Day	Float	26.60–125.45
Total bunker cost	USD	Float	1026.12–3,007,958.65

In the paper, we removed the outliers of several variables, including type, distance, oil price, speed, port days, and bunker cost, which are related to ship fuel consumption. The process of removing outliers involved using statistical methods such as box plots or normal distributions to identify outliers and then removing them from the dataset. This study carefully examined the data before removing outliers to ensure that the points were indeed unacceptable or incorrect data and not legitimate data points (Figure 3).

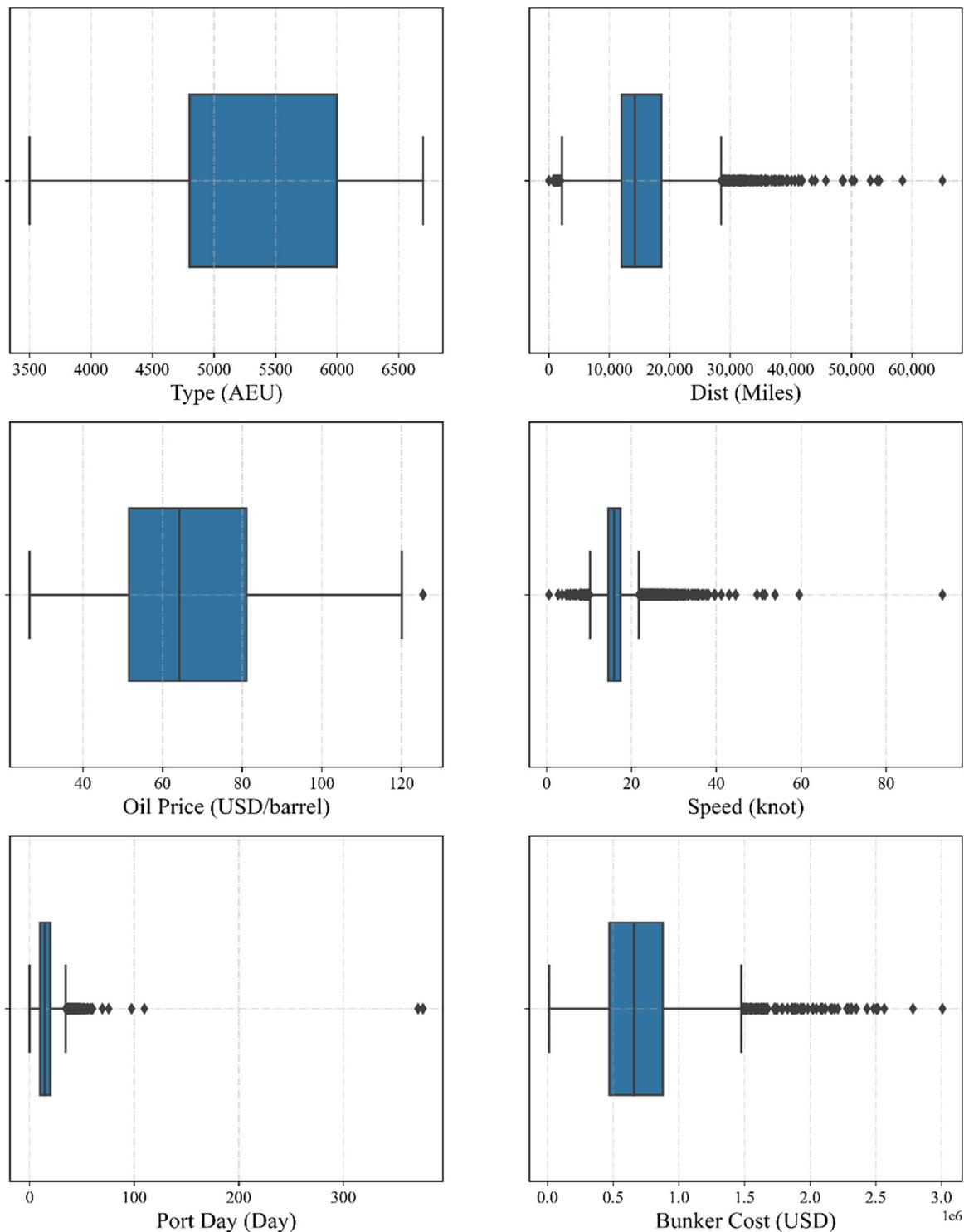


Figure 3. Outliers in the dataset.

In this paper, we used histograms to analyze the distribution of several variables, including type, distance, oil price, speed, and port days, which are related to ship fuel consumption. The histograms provided an overall picture of the data distribution, allowing us to identify any potential skewness, bimodality, or other patterns that may exist in the data. By examining the histograms, we were able to gain a better understanding of the data and make informed decisions regarding data cleaning, modeling, and other analytical techniques. For example, Figure 4 gives the frequency of occurrence of each value in the dataset.

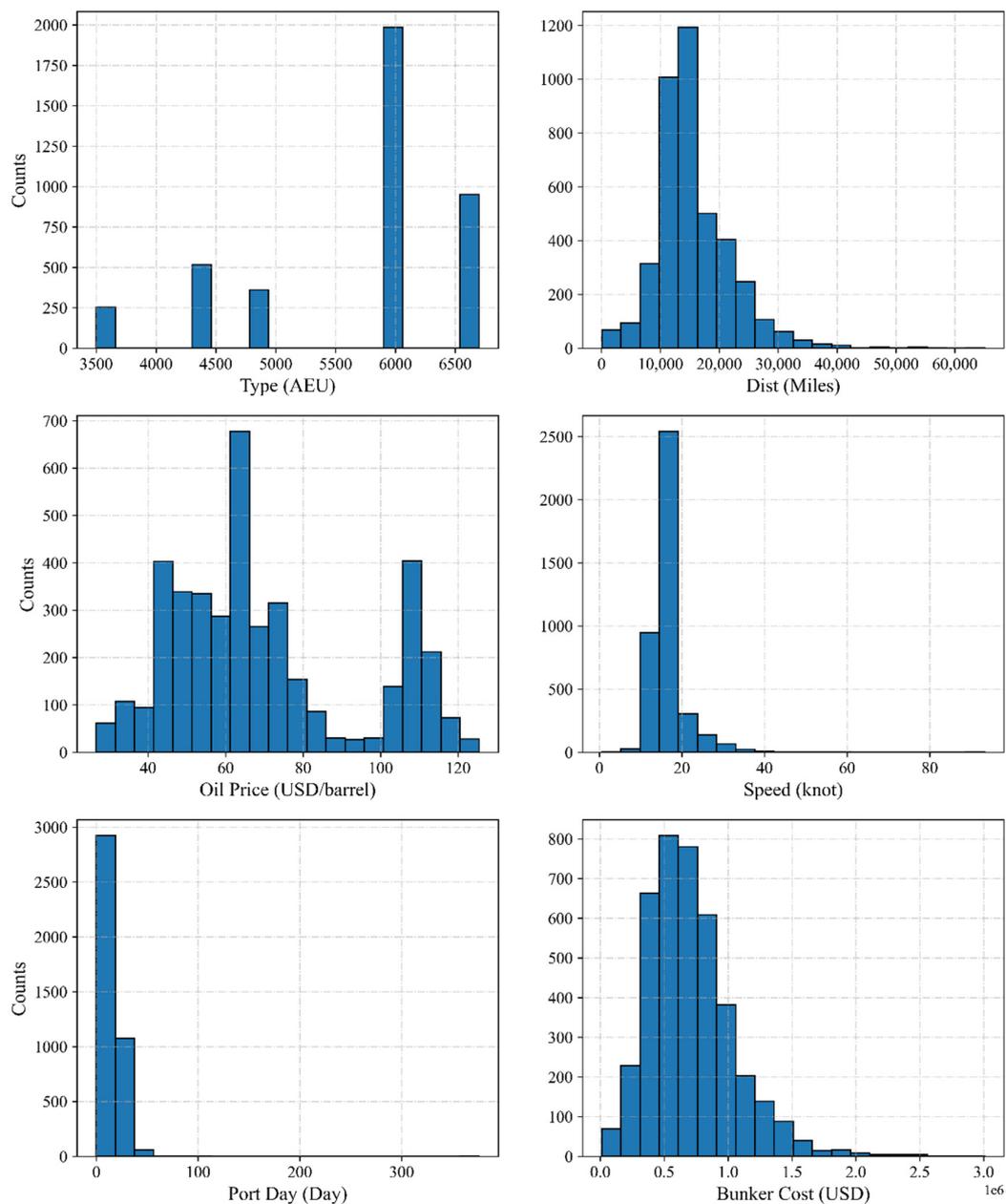


Figure 4. Histogram showing the frequency distribution of the data set.

3.5. Data Pre-Processing

The problem of data pre-processing is fundamental to data-driven modeling [64]. By pre-processing and removing data, we can identify and clean up data anomalies [36]. In this study, data pre-processing included examining data distributions and correlations. Determine their anomalous indicators. Remove all outlier data discovered by the joint operation. Ultimately, a clean and usable dataset is obtained.

A method known as feature scaling may be used to standardize the values of the independent variables that comprise a data collection within a certain range. To put it another way, feature scaling narrows the scope of variables so that we can make meaningful comparisons across sets of data. Unscaled data slow down the convergence process, which occurs during data pre-processing to deal with the magnitude or value of high variables. The following describes the normalization algorithms:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{14}$$

We split each dataset of the machine learning model into a training set and a test set. We picked 70% of the data items for the training set, leaving 30% for the test set. The above model provides a data-driven approach to fuel cost modeling and optimization that can provide an analytical approach and a reference for low fuel costs in maritime transport.

4. Results

4.1. Modeling

We obtained 16,189 valid data pieces from 2012 to 2021 following data cleaning. We partitioned the data in a 7:3 ratio between training and test sets. After normalizing the data, we used multiple regression algorithms to train on fuel costs.

The training set trains the models to make good predictions, and the test set tests the ability of the trained models. Then, we utilized a variety of algorithms (Table 4) to instruct the training set data. The average MSE and R² for each model after 100 training trials are displayed in Table 4. The results of the trained models on the test sets and the training sets are as follows (MSE and R² are evaluation indicators):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (\bar{y}_i - \hat{y}_i)^2} \tag{16}$$

y_i —Real values.

\hat{y}_i —Predicted values.

\bar{y}_i —Average.

Table 4. Comparison of model performance.

Models	Training Set	Test Set	Training Set R ²	Test Set R ²
Linear model	3.962×10^3	2.603×10^3	0.738	0.780
Random Forest	4.037×10^4	1.341×10^3	0.973	0.886
DT	0×10^{00}	3.127×10^3	1.000	0.735
SVM	2.874×10^3	1.797×10^3	0.810	0.832
KNN	1.952×10^3	1.797×10^3	0.871	0.848
Adaboost	3.622×10^3	2.830×10^3	0.760	0.760
GBRT	1.683×10^3	1.244×10^3	0.889	0.895
Bagging	4.755×10^4	1.454×10^3	0.969	0.877
ExtraTree	0×10^{00}	3.621×10^3	1.000	0.693
LASSO	1.511×10^2	1.202×10^2	0.000	0.018
MLP	3.068×10^3	1.873×10^3	0.797	0.841
SGD	8.377×10^3	6.312×10^3	0.446	0.466
XGBR	2.406×10^4	9.633×10^4	0.984	0.968
BP NN	2.861×10^3	3.649×10^3	0.837	0.744
RBF NN	2.738×10^4	3.128×10^3	0.844	0.780

Based on overall performance, XGBoost has the best performance. In addition, we discovered that the predictions made by the standard XGBoost model for the target variables were already very good and accurate. No redundant changes are required. Therefore, we used the XGBR model as the base model in the subsequent optimization of the fuel costs. Figures 5 and 6 show the performance in XGBR (where the x-axis represents the data encoding, as the data are arbitrary).

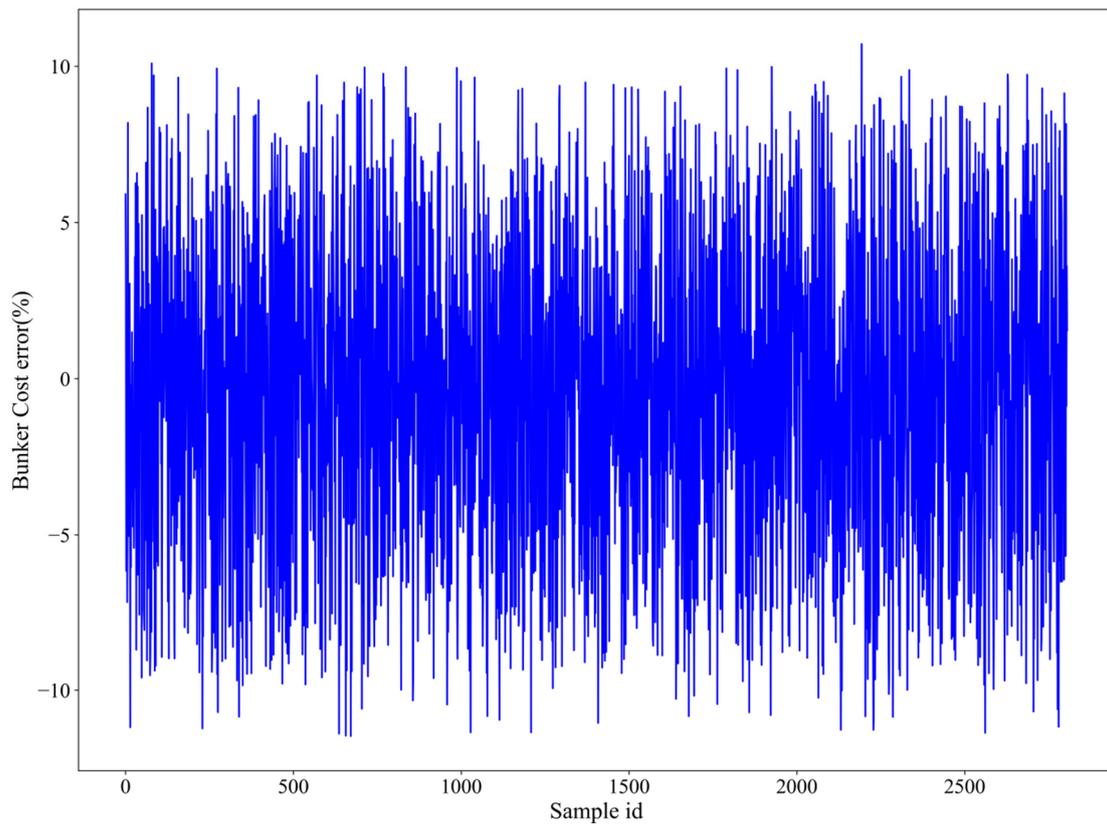


Figure 5. Relative error between predicted and tested values on the training set.

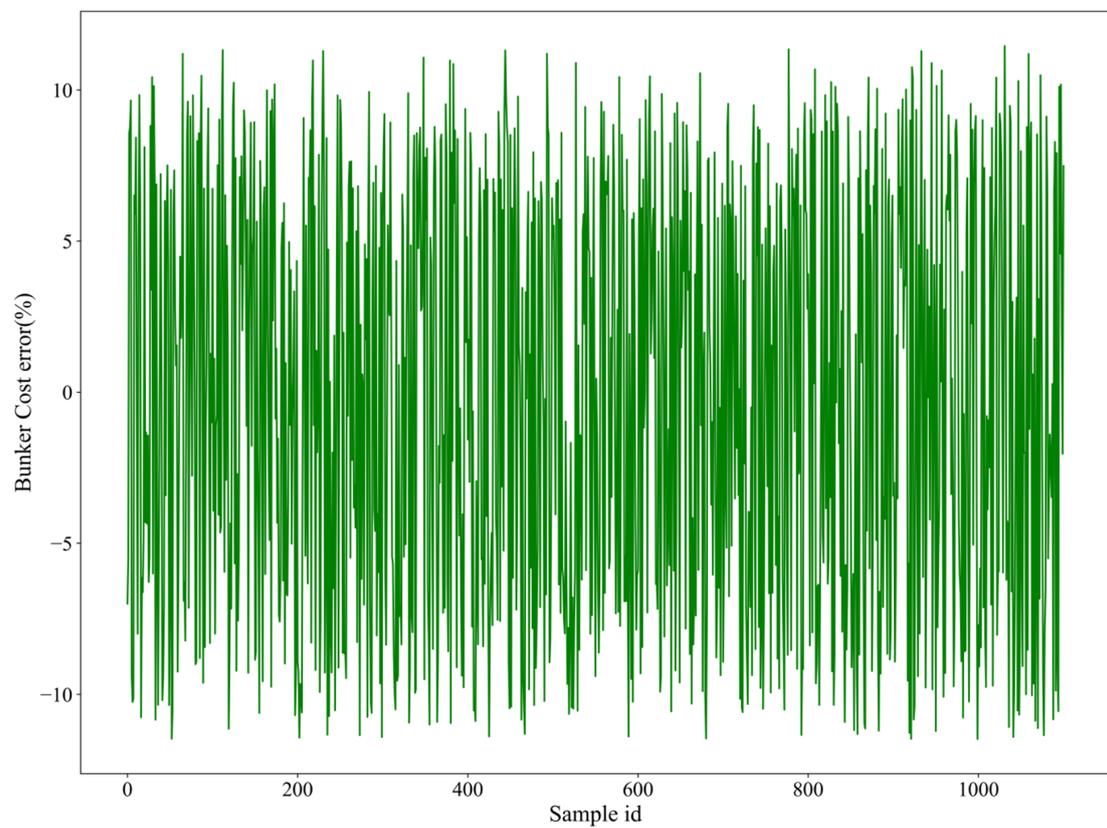


Figure 6. Relative error between predicted and tested values on the test set.

We also tested the uncertainty and robustness of the model. In the machine learning modeling process, it is common practice to divide the data into a training set and a test set. The test set is data that is independent of the training, is not involved in the training at all, and is used for the evaluation of the final model. During the training process, there is often an overfitting problem where the model can match the training data well but cannot predict the data outside the training set very well. Using the test data to adjust the model parameters at this point would be equivalent to knowing some of the information from the test data at the time of training, which would affect the accuracy of the final evaluation results. It is common practice to use a portion of the training data as validation data to evaluate the training effect of the model.

The validation data are taken from the training data but are not involved in the training, so that the model can be evaluated relatively objectively on how well it matches the data outside the training set. A common evaluation of models in validation data is cross-validation, also known as round-robin validation. It divides the original data into K groups (K-Fold) and makes a separate validation set for each subset of data, with the remaining K-1 subsets of data serving as the training set, resulting in K models. These K models are evaluated separately in the validation set, and the final error MSE (mean squared error) is summed and averaged to obtain the cross-validation error. Cross-validation ensures that the model correctly captures the pattern from the data, regardless of the interference from the data. The robustness of the model is indicated if the fit of all K models of the model is high, i.e., the model predicts very accurately with different combinations of training and test sets.

In this paper, a 10-fold split of the dataset was selected, and the R-squared of the training and test sets in the 10 groupings is shown in Figure 7, and the performance is excellent.

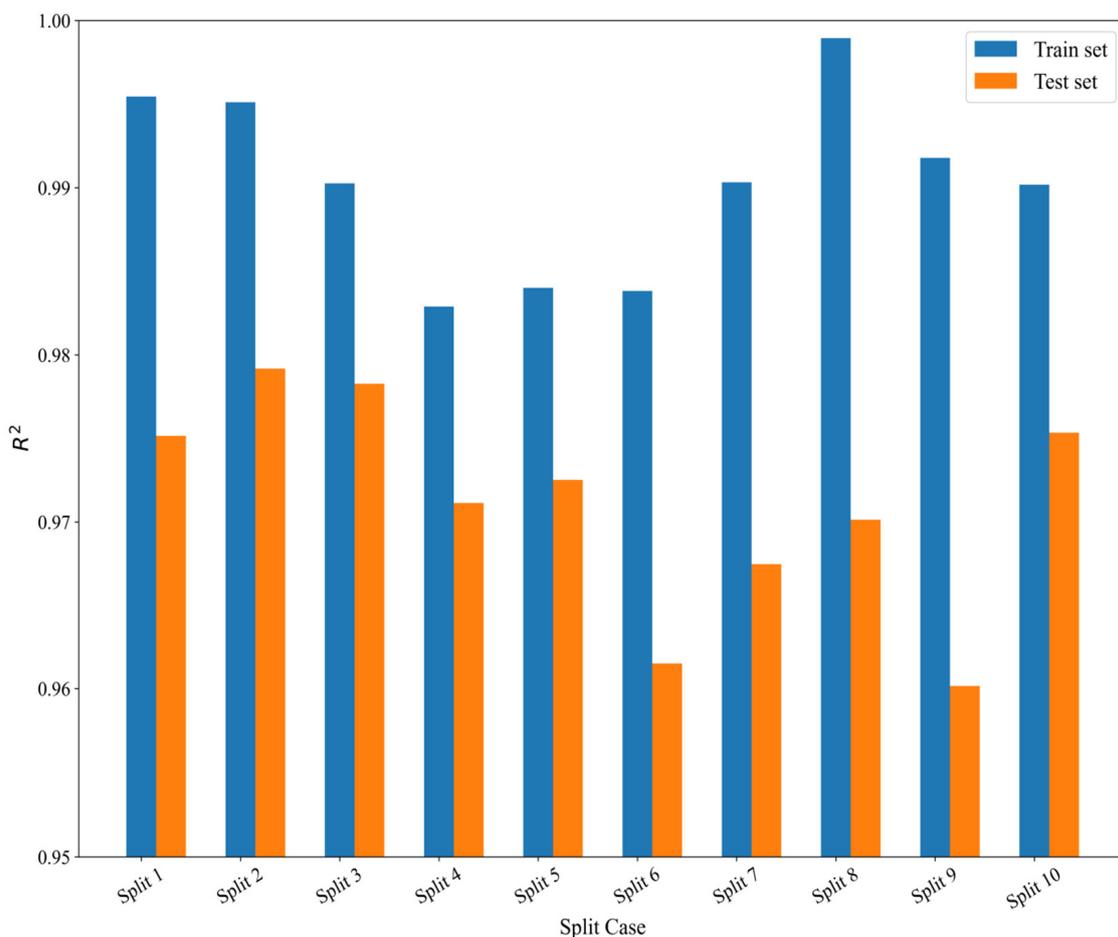


Figure 7. R-squared for the training and test sets in the 10 subgroups.

In this paper, the XGBoost model is used as a regressor to regress the data. There are two main sources of overall uncertainty in the model: data uncertainty (also known as chance uncertainty) and knowledge uncertainty (also known as cognitive uncertainty). Data uncertainty arises from the inherent complexity of the data, such as additive noise or overlapping classes. In these cases, the model knows that the input has attributes from more than one class or that the target is noisy. Importantly, it is not possible to reduce data uncertainty by collecting more training data. Knowledge uncertainty occurs when the model’s inputs come from regions where the training data are sparse or far from the training data. In these cases, the model knows very little about the region and may make mistakes. Unlike data uncertainty, knowledge uncertainty can be reduced by collecting more training data from a region that is poorly understood.

In order to estimate data uncertainty, it is necessary to use probabilistic regression models that predict the mean and variance. To this end, the following new loss function has been designed:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma(x_i)^2} \| y_i - f(x_i) \|^2 + \frac{1}{2} \log \sigma(x_i)^2$$

where sigma is data uncertainty.

Knowledge uncertainty, on the other hand, can be obtained by measuring the mean variance between multiple models. Different pruning parameters were used for XGBoost to design multiple models and then calculate their knowledge uncertainty.

Particularly, as depicted in Figure 8, the model for the training set’s uncertainty is as follows:

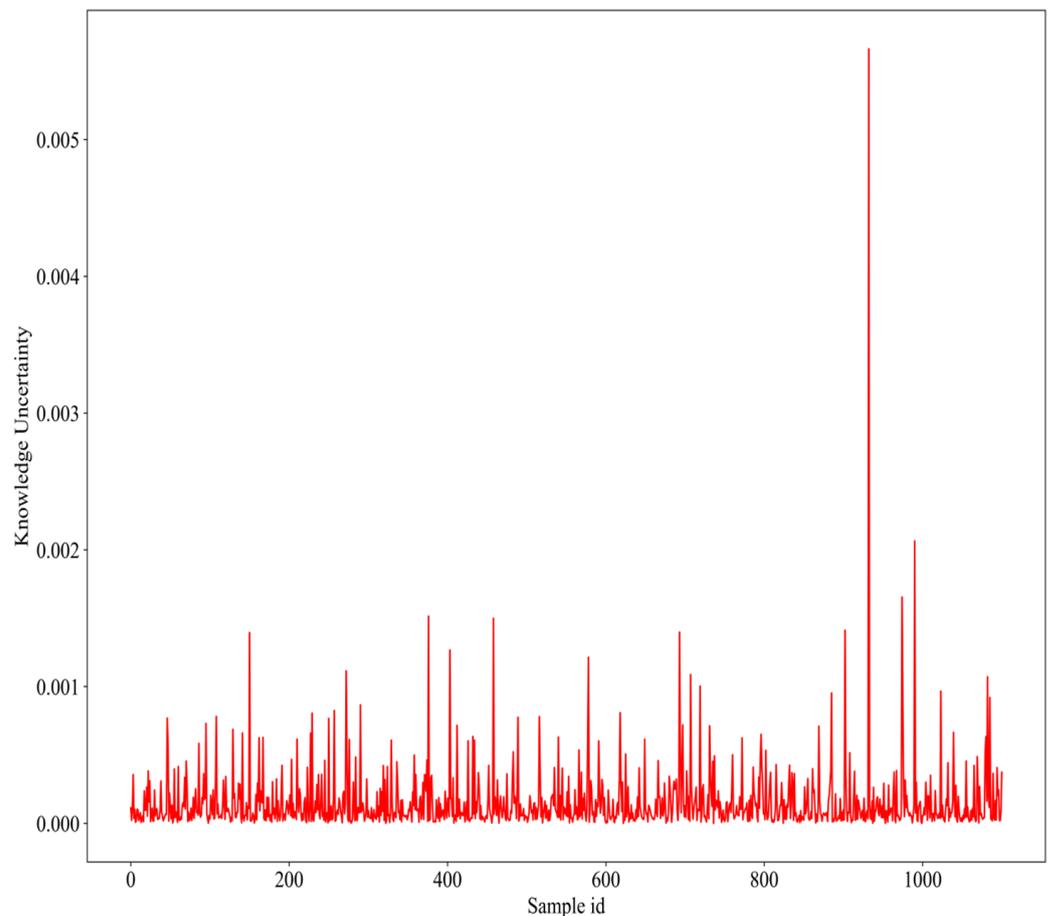


Figure 8. Cont.

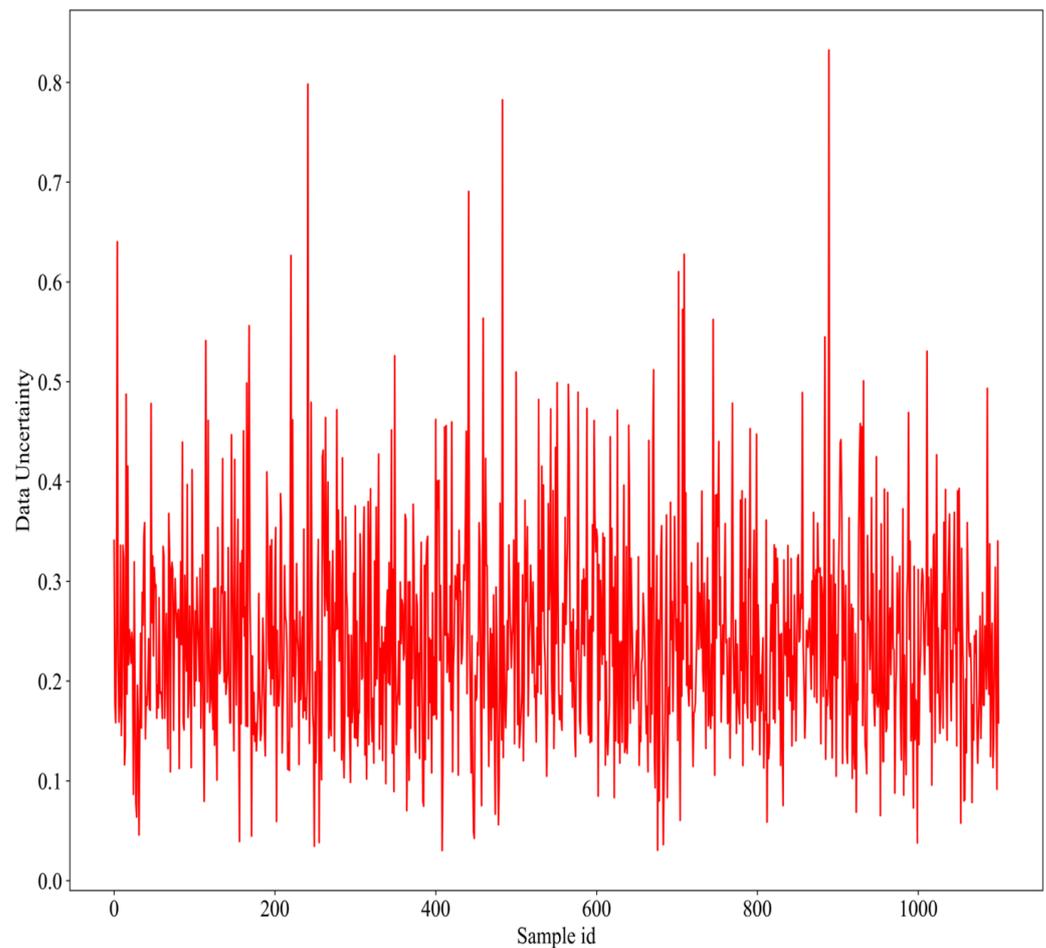


Figure 8. Uncertainty in the model.

4.2. Optimization

This paper focused on the optimal fuel cost problem with speed as the independent variable. That is, to find the optimal speed for a fixed type, distance, port day, and oil price. Particle swarm optimization methods are global optimization algorithms with advantages; we chose the PSO algorithm as the optimization algorithm. To further optimize the fuel cost, optimize the combination of variables, and lower the fuel cost, we selected the XGBR model as the mathematical basis for the particle swarm method.

4.2.1. Optimization Process

Using the trained XGBR model as the basic mathematical model for optimization, we took type, distance, port day, and oil price as constant values to find the optimal speed. For the three routes, ASNW, ASAG, and ASEU, and the two most conventional ship types, we conducted six scenarios with an oil price of 100 and a port day of 20 to select the optimal speed configuration, as shown in Table 5.

Table 5. Optimization conditions.

Case	Routes	Type	Distance	Oil Price	Port Day
1	ASNW	6700	12,000	100	20
2	ASNW	6000	12,000		
3	ASAG	6700	15,000		
4	ASAG	6000	15,000		
5	ASEU	6700	14,500		
6	ASEU	6000	14,500		

We performed a particle swarm search for the six cases above to find the optimal speed to minimize the bunker cost. First, we chose a population of 100 particles, with each particle position representing speed. This speed variable is limited in range during shipping, as shown in Table 6.

Table 6. Speed range of process parameters.

Variables	Minimum Value	Maximum Value
Speed	0	50

The maximum and minimum velocities may be used to calculate the accuracy of the region between the current and ideal positions. If the absolute value of the maximum velocity (or minimum velocity) is excessively high, the particles may pass through the target region because the accumulated inertial velocity is excessively high. This scenario inhibits effective searching for the global optimal solution. However, if the absolute value of the maximum velocity (or minimum velocity) is too small, the particles cannot quickly focus on the current global optimal solution and search its neighborhood effectively. In this situation, the particles easily fall into local extremes that one cannot increase. The velocity of each particle in the corresponding dimension in the optimization process is limited to the range of $[-\bar{v}, \bar{v}]$, where \bar{v} is the 2% multiplication of the maximum minus lowest value of the variable range in the given dimension.

First, we randomly generated 100 initial positions within the process parameters. To determine if a particle’s location is ideal, we substituted the XGBR model’s output (fuel cost) for the fitness value; the lower the fitness value, the better the position. This approach is the opposite of the traditional criterion of greater fitness being better. These coordinates are then normalized using the XGBR model used to determine fitness values and sent into the neural network as input to complete the calculation. The program then repeats the processes throughout the iterations in accordance with the speed and position update formulas, constantly updating the optimal positions experienced by each particle and the global optimal positions until the maximum number of iterations is reached.

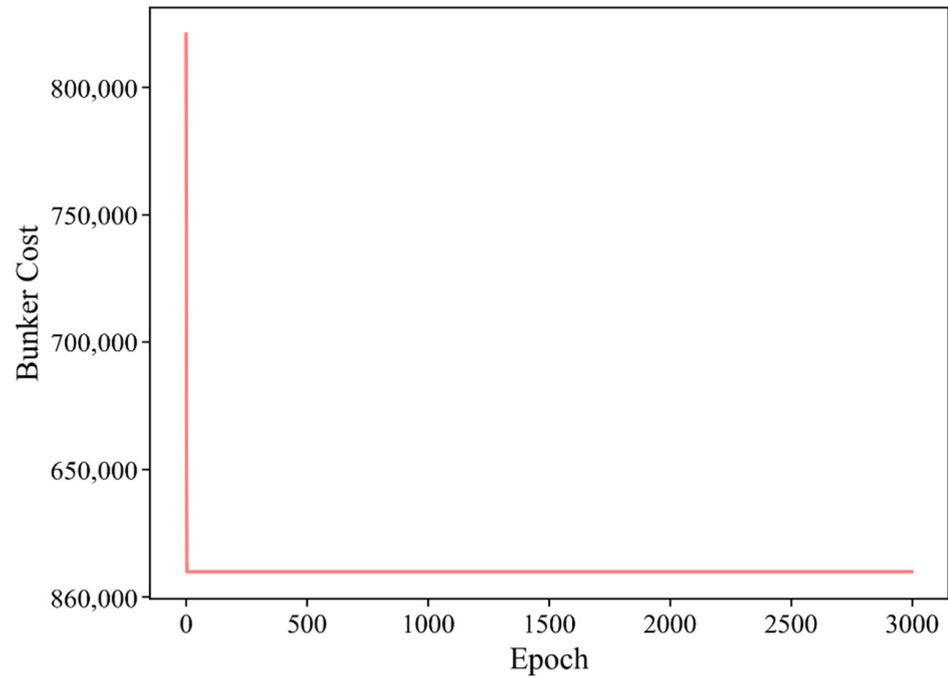
Specifically, a larger number of iterations enables the algorithm to investigate the solution space more thoroughly and, consequently, increases the likelihood of locating the global optimal solution. Consequently, a reduced maximum number of iterations decreases the likelihood that the algorithm will discover the global optimal solution; it may cause the optimization process to stop before the particle has experienced the optimal position. In this paper, we capped the total number of iterations at 3000, and empirical optimization results back up this decision.

4.2.2. ASNW Route Optimization

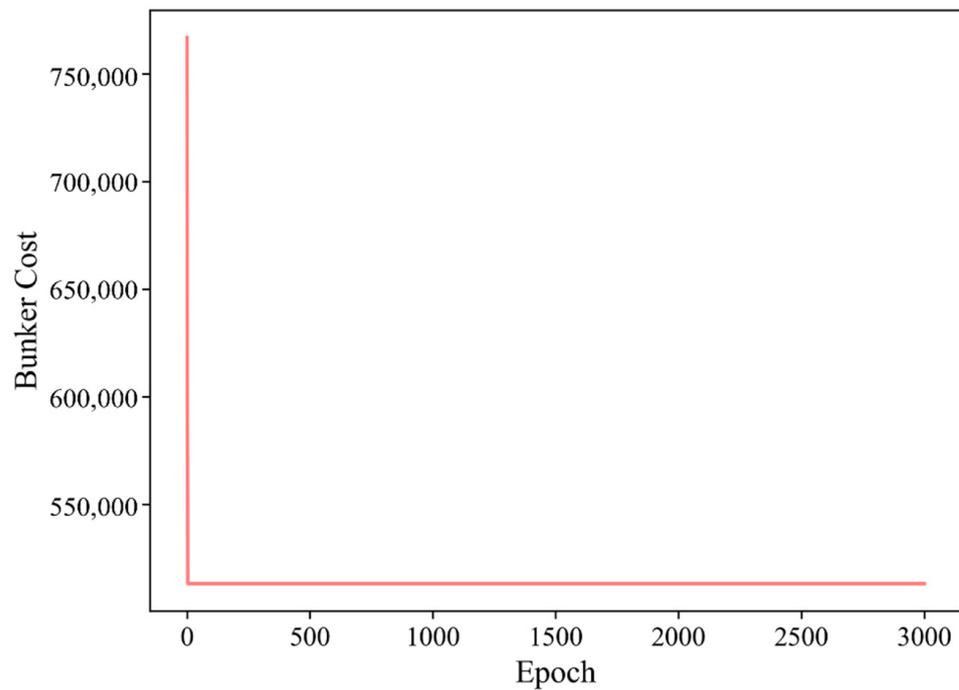
Table 5 (above) shows the optimization of Cases 1 and 2 on the ASNW route. Figure 9 shows the change in fitness function during this optimization process.

Figure 9 also provides the optimal fuel cost curves for the PSO algorithm optimization process in two cases. Clearly, as the number of iterations increases, the optimal fuel cost decreases significantly in all three cases. In the early stages of optimization, the optimal fuel cost decreases quickly; as the number of iterations increases, the optimal fuel cost decreases more slowly until the late stages of optimization, when it stabilizes and reaches the optimal solution. This curve also justifies the choice of the maximum number of iterations, as the minimum value stabilizes with increasing numbers of iterations in the later stages of the optimization. Table 7 summarizes the results of the optimization. The optimization process presented in Tables 7–9 was conducted with a rigorous approach. Multiple rounds of validation of the modeling results and optimization errors in practice were carried out. The results were obtained after 100 rounds of optimization, ensuring a thorough analysis. The mean value has been calculated and is presented as “mean +/- standard deviation” to account for any potential errors in the data. This value accurately reflects the optimized results and can be considered reliable. Based on these findings, we can conclude that the

results obtained are trustworthy and can serve as a solid foundation for further research and analysis.



(a)



(b)

Figure 9. Particle swarm optimization process curve. (a) Case 1; (b) Case 2.

Table 7. Case 1; Case 2 optimization results.

Case	Best Speed	Best Bunker Cost
1	35.37 ± 1.02	609,739.25 ± 9024.13
2	34.30 ± 0.85	513,247.34 ± 5799.60

Table 8. Case 3; Case 4 optimization results.

Case	Best Speed	Best Bunker Cost
3	34.80 ± 1.14	844,674.70 ± 6314.23
4	35.14 ± 1.03	789,084.20 ± 6067.58

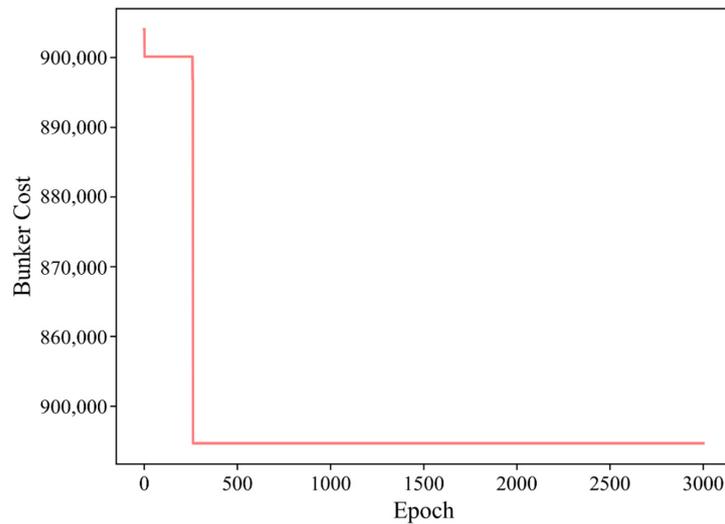
Table 9. Case 5; Case 6 range of process parameters.

Case	Best Speed	Best Bunker Cost
5	33.99 ± 0.98	846,517.50 ± 8893.67
6	35.37 ± 1.25	790,927.06 ± 4529.43

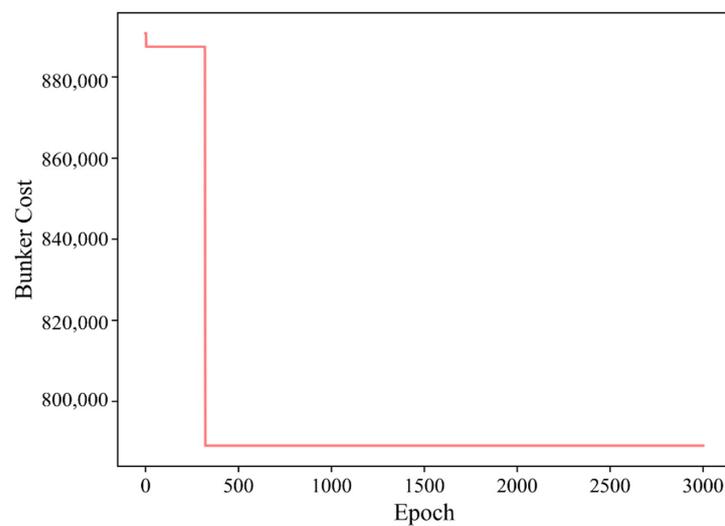
Thus, for the ASNW route, about 35 is optimal for the average speed (range 0–50), and the optimal cost for a vessel of Type 6000 is less than that of Type 6700.

4.2.3. Optimization of the ASAG Route

Table 5 (above) details the optimization of the ASAG route for Cases 3 and 4. In addition, Figure 10 illustrates the change in fitness function during the optimization process.



(a)



(b)

Figure 10. Particle swarm optimization process curve. (a) Case 3; (b) Case 4.

In addition, Figure 10 depicts the variation curves of the optimal fuel cost during the optimization of the PSO algorithm for the two cases, while Table 8 displays the optimization outcomes.

It appears that 35 is optimal for the average speed (range 0–50) for the ASAG route, and the optimal cost for a Type 6000 vessel is less than for a Type 6700.

4.2.4. Optimization of the ASEU Route

As with the other routes, Table 5 summarizes the optimization on the ASEU route in Cases 5 and 6, while Figure 11 illustrates the optimization process for the change in the fitness function.

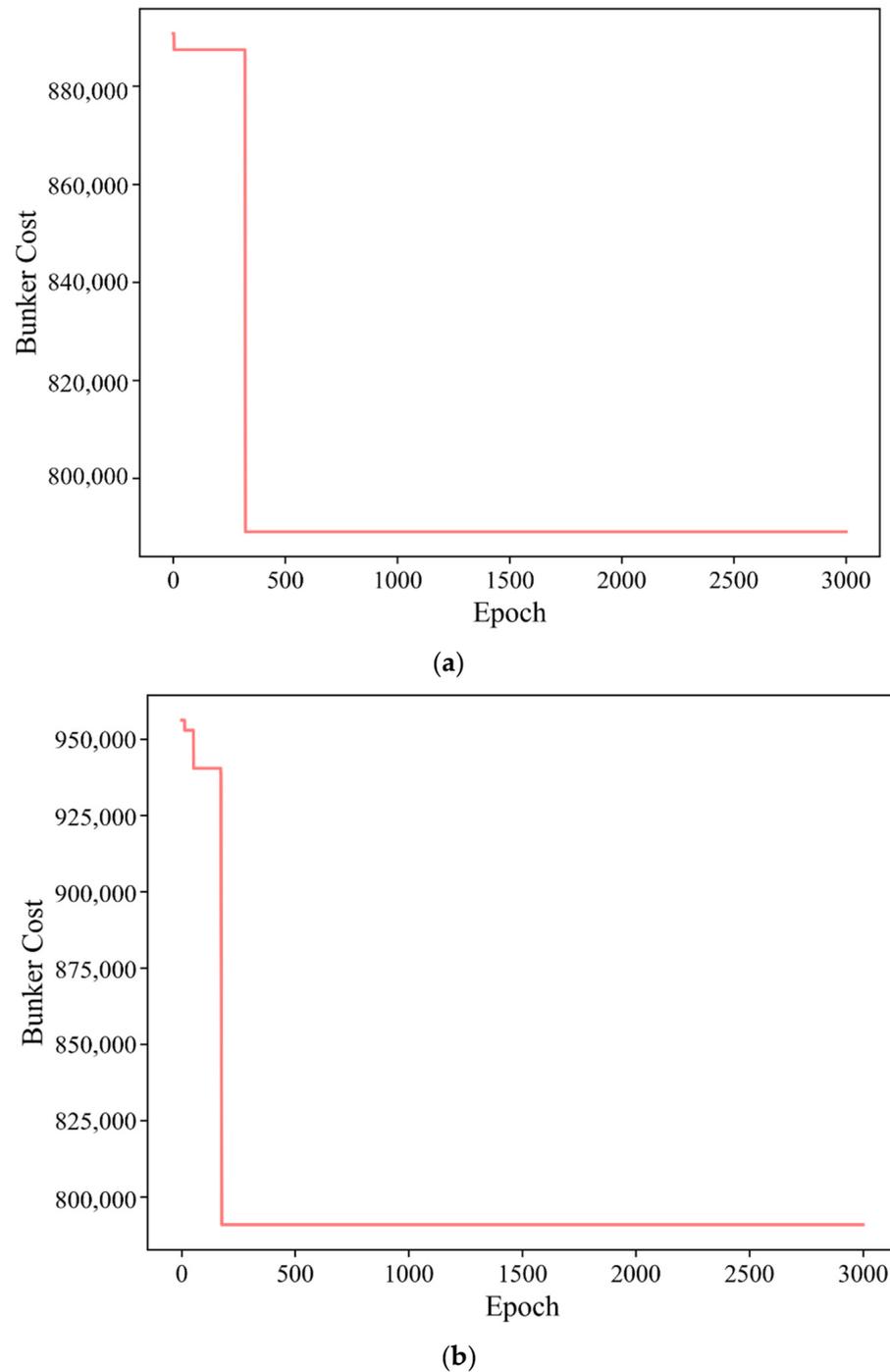


Figure 11. Particle swarm optimization process curve. (a) Case 5; (b) Case 6.

Figure 11 displays the ideal fuel cost variation curves generated by the PSO method during optimization for the two scenarios. The outcomes of the optimization are shown in Table 9.

The best average speed (range 0–50) for the ASEU routes appears to be around 33, and the optimal cost for a Type 6000 vessel is less than for a Type 6700. Therefore, for the same vessel type, the cost of the ASNW route is much lower than the ASAG and ASEU routes, while the optimal cost of the ASAG and ASEU routes is approximately the same.

The above model provides a data-driven approach to fuel cost modeling and optimization that can provide an analytical approach and a reference for low fuel costs in maritime transport.

5. Conclusions and Discussion

Monitoring, predicting, and optimizing ship fuel consumption are crucial components of ship energy sustainability management. In this investigation, we created an integration model for predicting fuel consumption using deep learning. We also provided an optimization model for this purpose using data from Korea's largest pure car and truck shipping company for the last ten years (2012–2021). Specifically, in the first stage, we pre-processed and thoroughly analyzed ten years of ship operation data from pure car and truck shipping companies to determine the integrated fuel consumption model for ships at different speeds, vessel types, port days, distances, oil prices, etc. This process also resulted in some improvement in the predictive performance of all models.

Second, we examined some of the most well-known machine learning models, including random forests, support vector machines, and decision trees. The comparison results revealed the factors that significantly affect the results of the models and which machine learning algorithms are robust. The findings prove that XGBoost is the most trustworthy model (R^2 value of 0.97) for estimating pure car/truck ship energy consumption and logistics costs in the shipping industry. Subsequently, we proposed a particle swarm optimization model in the second stage. By calculating the optimal sailing speed for each segment of the voyage, we can significantly cut down on the ship's CO_2 emissions and fuel usage throughout the course of the whole trip. In addition, we analyzed the degree of influence of input characteristics on total fuel consumption.

We further noted that speed variations could significantly affect the results of our optimization model. Such prediction and optimization models can help pure car and truck ship operators and shipping companies make more cost-effective decisions (e.g., estimating order fuel consumption, making reasonable offers, providing better digital solutions, etc.) and establish and improve their corporate energy metering systems.

The following are the main findings of this study: First, a multi-source dataset of sailing status, speed variation, and segmentation information can assist marine vessel fuel consumption fitting analysis. In this study, we compared the ability of different models to calculate real-time fuel consumption rates from data. We found significant improvement in the performance of each prediction model after adding other speed data and time in port days, etc., as input feature variables and using hyperparameter optimization (the R^2 for the XGBoost model is 0.97, which is the maximum value). Second, speed optimization can effectively increase the energy efficacy of ships and reduce bunker costs by a considerable amount. Third, the impact of route and days in port on ship petroleum consumption varies among the tested algorithms, as does the effect on predicted performance. In the end, we studied and modeled the vessel's actual fuel usage, then optimized fuel consumption and overall cost throughout the whole range. We determined that the proposed XGBoost and PSO models have good accuracy and robustness.

Our fuel consumption studies have important implications for the study of global climate change mitigation. A significant tool for attaining goals such as cost reduction and greenhouse gas reduction is the employment of proper predictive algorithms to assess a ship's fuel usage before (or during) a voyage [28]. Scholars are addressing the increase in fuel consumption in the transport sector, including the shipping transport industry [26].

Furthermore, the fuel consumption of existing liner carriers has a direct impact on the operating expenses of stakeholders and the rise in greenhouse gas emissions.

Due to the need for sustainable technology in maritime transport, adaptive machine learning tools can provide the greatest efficiency, sustainability, and reduced operating costs [65]. In the last decade, researchers have conducted various studies on machine learning in different industries (telecommunications, computing, aviation, railways, machinery, etc.). However, few studies exist on maritime fuel consumption [65], and only a few studies have used non-classical methods to estimate the performance of ships in maritime navigation [28]. Therefore, improving energy efficiency is still an open topic. This study, therefore, enriches the machine learning literature in the maritime fuel area.

Our work introduces a novel XGBoost fuel consumption ensemble model and particle swarm fuel consumption optimization model in an effort to increase the energy efficiency of transport companies. Although it is virtually impossible to obtain infallible results due to the occurrence of dynamically variable conditions in real time, the genuine and extensive data used in this study can more accurately reflect the actual operation of shipping companies than experimental data. Therefore, we believe this study's research results are close to their actual value. This application means that vessels can effectively change their parameters, such as speed and port dates, to respond to the situation and help car-and-truck-only shipping companies make more economical and environmentally friendly decisions. Hence, the model given in this study may be used as the foundation for a pure car/truck ship energy efficiency management program and then expanded to include real-time monitoring of ship energy efficiency and the identification of anomalous fuel usage. This study's two-stage model is also useful for predicting and optimizing ship fuel consumption for other ship types, ships of the same type, ships in different basins, and ships operated by different ship proprietors. This is because these factors do not alter the ship's fundamental physical characteristics. Moreover, this research brings PC/TC shipping businesses closer to their cost-cutting and environmentally friendly goals.

6. Limitations and Future Research

We have made every effort to acquire and analyze experimental data and have proposed a two-step method for determining the optimal ship speed to reduce fuel consumption. However, we have only evaluated the models provided in this study against region-specific data. We acknowledge that there are some limitations to this study. Consequently, future research should evaluate the impact of additional constraints, such as weather, sea conditions, propeller fouling, and marine organisms. We can add more data elements from various sources, such as ocean and weather data from weather forecasting websites, and use larger samples to analyze and affirm the robustness and accuracy of the proposed model. In addition, we acknowledge that every vessel has its own distinct characteristics and operational conditions, so our model may require modifications and optimizations to better adapt to various ship situations. In the following paper, we intend to conduct additional experiments to evaluate and validate the proposed model and report the results in order to confirm its accuracy and dependability. We will provide additional explanations and analyses to enhance the credibility and professionalism of the research findings.

Author Contributions: Conceptualization, methodology, writing—original draft, review and editing, software, formal analysis, data curation, M.S. and S.C.; investigation, resources, funding acquisition, review and editing, software, formal analysis, data curation, writing—review and editing, S.-H.B. and Z.S.; review and editing, K.-S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Thank you to the editor and two reviewers for their insightful comments.

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

Appendix A

Table A1 shows the full names and abbreviations of the terms used in this paper.

Table A1. Full name and abbreviation.

Full Name	Abbreviation
Pure car and truck shipping company	PCTC
Black box model	BBM
Machine learning	ML
Particle swarm optimization	PSO
International Maritime Organization	IMO
Greenhouse Gas	GHG
Multiple linear regression	MLR
Artificial neural network	ANN
Deep reinforcement learning	DRL
Emission control areas	ECAs
Mean absolute percentage error	MAPE
Computational Fluid Dynamics	CFD
Extreme gradient boosting	XGBoost
Gradient Boosting Decision Tree	GBDT
Pure car carrier	PCC
Accent Equivalent Unit	AEU
Marine Gas Oil	MGO
Least absolute shrinkage and selection operator	LASSO
Ridge regression	RIDGE
Support vector regression	SVR
K Nearest Neighbors	KNN
Regressive Function	RF
Decision Tree	DT
Asia North West America	ASNW
Asia Arabian Gulf	ASAG
Asia Europe	ASEU
380 Centistoke High Sulfur Fuel Oil	380 CST HSFO
Marine Gas Oil	MGO
Distance	Dist
Long Short–Term Memory	LSTM

References

- Zhao, J.; Yang, L. A bi-objective model for vessel emergency maintenance under a condition-based maintenance strategy. *Simulation* **2018**, *94*, 609–624. [[CrossRef](#)]
- Zhang, J.; Teixeira, Á.P.; Soares, C.G.; Yan, X. Quantitative assessment of collision risk influence factors in the Tianjin port. *Saf. Sci.* **2018**, *110*, 363–371. [[CrossRef](#)]
- Hu, Z.; Zhou, T.; Zhen, R.; Jin, Y.; Li, X.; Osman, M.T. A two-step strategy for fuel consumption prediction and optimization of ocean-going ships. *Ocean Eng.* **2022**, *249*, 110904. [[CrossRef](#)]
- Bagoulla, C.; Guillotreau, P. Maritime transport in the French economy and its impact on air pollution: An input-output analysis. *Mar. Policy* **2020**, *116*, 103818. [[CrossRef](#)]
- Teixeira, A.P.; Soares, C.G. Reliability analysis of a tanker subjected to combined sea states. *Probabilistic Eng. Mech.* **2009**, *24*, 493–503. [[CrossRef](#)]
- Li, X.; Du, Y.; Chen, Y.; Nguyen, S.; Zhang, W.; Schönborn, A.; Sun, Z. Data fusion and machine learning for ship fuel efficiency modeling: Part I–Voyage report data and meteorological data. *Commun. Transp. Res.* **2022**, *2*, 100074. [[CrossRef](#)]
- Yang, L.; Chen, G.; Rytter, N.G.M.; Zhao, J.; Yang, D. A genetic algorithm-based grey-box model for ship fuel consumption prediction towards sustainable shipping. *Ann. Oper. Res.* **2019**, 1–27. [[CrossRef](#)]
- Yuan, Z.; Liu, J.; Zhang, Q.; Liu, Y.; Yuan, Y.; Li, Z. Prediction and optimisation of fuel consumption for inland ships considering real-time status and environmental factors. *Ocean Eng.* **2021**, *221*, 108530. [[CrossRef](#)]
- Yan, R.; Wang, S.; Du, Y. Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *138*, 101930. [[CrossRef](#)]

10. Hu, Z.; Zhou, T.; Osman, M.T.; Li, X.; Jin, Y.; Zhen, R. A novel hybrid fuel consumption prediction model for ocean-going container ships based on sensor data. *J. Mar. Sci. Eng.* **2021**, *9*, 449. [\[CrossRef\]](#)
11. Işıklı, E.; Aydın, N.; Bilgili, L.; Toprak, A. Estimating fuel consumption in maritime transport. *J. Clean. Prod.* **2020**, *275*, 124142. [\[CrossRef\]](#)
12. Schramm, H.J. A cliometric approach to market structure and market conduct in the car carrier industry. *Case Stud. Transp. Policy* **2020**, *8*, 394–402. [\[CrossRef\]](#)
13. Banawan, A.; Mosleh, M.; Seddiek, I. Prediction of the fuel saving and emissions reduction by decreasing speed of a catamaran. *J. Mar. Eng. Technol.* **2013**, *12*, 40–48.
14. Yan, R.; Wang, S.; Psaraftis, H.N. Data analytics for fuel consumption management in maritime transportation: Status and perspectives. *Transp. Res. Part E Logist. Transp. Rev.* **2021**, *155*, 102489. [\[CrossRef\]](#)
15. Leifsson, L.P.; Sævarsdóttir, H.; Sigurðsson, S.P.; Vésteinsson, A. Grey-box modeling of an ocean vessel for operational optimization. *Simul. Model. Pract. Theory* **2008**, *16*, 923–932. [\[CrossRef\]](#)
16. Adland, R.; Cariou, P.; Wolff, F.-C. Optimal ship speed and the cubic law revisited: Empirical evidence from an oil tanker fleet. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *140*, 101972. [\[CrossRef\]](#)
17. Wang, S.; Meng, Q. Sailing speed optimization for container ships in a liner shipping network. *Transp. Res. Part E Logist. Transp. Rev.* **2012**, *48*, 701–714. [\[CrossRef\]](#)
18. Kee, K.-K.; Simon, B.-Y.L.; Renco, K.-H. Artificial neural network back-propagation based decision support system for ship fuel consumption prediction. In Proceedings of the 5th IET International Conference on Clean Energy and Technology (CEAT2018), Kuala Lumpur, Malaysia, 5–6 September 2018.
19. Adland, R.; Cariou, P.; Jia, H.; Wolff, F.-C. The energy efficiency effects of periodic ship hull cleaning. *J. Clean. Prod.* **2018**, *178*, 1–13. [\[CrossRef\]](#)
20. Meng, Q.; Du, Y.; Wang, Y. Shipping log data based container ship fuel efficiency modeling. *Transp. Res. Part B Methodol.* **2016**, *83*, 207–229. [\[CrossRef\]](#)
21. Zaman, I.; Pazouki, K.; Norman, R.; Younessi, S.; Coleman, S. Development of automatic mode detection system by implementing the statistical analysis of ship data to monitor the performance. *Int. J. Marit. Eng.* **2017**, *159*, 225–235. [\[CrossRef\]](#)
22. Bialystocki, N.; Konovessis, D. On the estimation of ship's fuel consumption and speed curve: A statistical approach. *J. Ocean Eng. Sci.* **2016**, *1*, 157–166. [\[CrossRef\]](#)
23. Goldstein, H. *Multilevel Statistical Models*; John Wiley Sons: Chichester, UK, 2011.
24. Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; Wasserman, W. *Applied Linear Statistical Models*; John Wiley & Sons: Hoboken, NJ, USA, 1996.
25. Kim, Y.-R.; Jung, M.; Park, J.-B. Development of a fuel consumption prediction model based on machine learning using ship in-service data. *J. Mar. Sci. Eng.* **2021**, *9*, 137. [\[CrossRef\]](#)
26. Tran, T.A. Comparative analysis on the fuel consumption prediction model for bulk carriers from ship launching to current states based on sea trial data and machine learning technique. *J. Ocean Eng. Sci.* **2021**, *6*, 317–339. [\[CrossRef\]](#)
27. Ahlgren, F.; Thern, M. Auto machine learning for predicting ship fuel consumption. In Proceedings of the ECOS 2018—The 31st International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, Guimarães, Portugal, 17–21 June 2018.
28. Farag, Y.B.; Ölçer, A.I. The development of a ship performance model in varying operating conditions based on ANN and regression techniques. *Ocean Eng.* **2020**, *198*, 106972. [\[CrossRef\]](#)
29. Du, Y.; Meng, Q.; Wang, S.; Kuang, H. Two-phase optimal solutions for ship speed and trim optimization over a voyage using voyage report data. *Transp. Res. Part B Methodol.* **2019**, *122*, 88–114. [\[CrossRef\]](#)
30. Gkerekos, C.; Lazakis, I.; Theotokatos, G. Machine learning models for predicting ship main engine fuel oil consumption: A comparative study. *Ocean Eng.* **2019**, *188*, 106282. [\[CrossRef\]](#)
31. Petersen, J.P.; Winther, O.; Jacobsen, D.J. A machine-learning approach to predict main energy consumption under realistic operational conditions. *Ship Technol. Res.* **2012**, *59*, 64–72. [\[CrossRef\]](#)
32. Uyanık, T.; Karatug, Ç.; Arslanoğlu, Y. Machine learning approach to ship fuel consumption: A case of container vessel. *Transp. Res. Part D Transp. Environ.* **2020**, *84*, 102389. [\[CrossRef\]](#)
33. Gkerekos, C.; Lazakis, I.; Papageorgiou, S. Leveraging big data for fuel oil consumption modelling. In Proceedings of the 17th Conference on Computer and IT Applications in the Maritime Industries, Pavone, Italy, 14–16 May 2018.
34. Le, L.T.; Lee, G.; Park, K.-S.; Kim, H. Neural network-based fuel consumption estimation for container ships in Republic of Korea. *Marit. Policy Manag.* **2020**, *47*, 615–632. [\[CrossRef\]](#)
35. Coraddu, A.; Oneto, L.; Baldi, F.; Anguita, D. Vessels fuel consumption forecast and trim optimisation: A data analytics perspective. *Ocean Eng.* **2017**, *130*, 351–370. [\[CrossRef\]](#)
36. Zhou, T.; Hu, Q.; Hu, Z.; Zhen, R. An adaptive hyper parameter tuning model for ship fuel consumption prediction under complex maritime environments. *J. Ocean Eng. Sci.* **2022**, *7*, 255–263. [\[CrossRef\]](#)
37. Armstrong, V.N. Vessel optimisation for low carbon shipping. *Ocean Eng.* **2013**, *73*, 195–207. [\[CrossRef\]](#)
38. Smith, T.; Parker, S.; Rehmatulla, N. On the speed of ships. In Proceedings of the International Conference on Technologies, Operations, Logistics and Modelling for Low Carbon Shipping, LCS2011, Glasgow, UK, 22 June 2011; University of Strathclyde: Glasgow, UK, 2011; pp. 22–24.

39. Fagerholt, K.; Laporte, G.; Norstad, I. Reducing fuel emissions by optimizing speed on shipping routes. *J. Oper. Res. Soc.* **2010**, *61*, 523–529. [[CrossRef](#)]
40. Li, X.; Sun, B.; Zhao, Q.; Li, Y.; Shen, Z.; Du, W.; Xu, N. Model of speed optimization of oil tanker with irregular winds and waves for given route. *Ocean Eng.* **2018**, *164*, 628–639. [[CrossRef](#)]
41. Psaraftis, H.N.; Kontovas, C.A. Ship speed optimization: Concepts, models and combined speed-routing scenarios. *Transp. Res. Part C Emerg. Technol.* **2014**, *44*, 52–69. [[CrossRef](#)]
42. Wen, M.; Pacino, D.; Kontovas, C.; Psaraftis, H. A multiple ship routing and speed optimization problem under time, cost and environmental objectives. *Transp. Res. Part D Transp. Environ.* **2017**, *52*, 303–321. [[CrossRef](#)]
43. Capezza, C.; Coleman, S.; Lepore, A.; Palumbo, B.; Vitiello, L. Ship fuel consumption monitoring and fault detection via partial least squares and control charts of navigation data. *Transp. Res. Part D Transp. Environ.* **2019**, *67*, 375–387. [[CrossRef](#)]
44. Bui-Duy, L.; Vu-Thi-Minh, N. Utilization of a deep learning-based fuel consumption model in choosing a liner shipping route for container ships in Asia. *Asian J. Shipp. Logist.* **2021**, *37*, 1–11. [[CrossRef](#)]
45. Chen, L.; Yip, T.L.; Mou, J. Provision of Emission Control Area and the impact on shipping route choice and ship emissions. *Transp. Res. Part D Transp. Environ.* **2018**, *58*, 280–291. [[CrossRef](#)]
46. Meng, Q.; Wang, S. Optimal operating strategy for a long-haul liner service route. *Eur. J. Oper. Res.* **2011**, *215*, 105–114. [[CrossRef](#)]
47. Davies, D.; Jindal-Snape, D.; Digby, R.; Howe, A.; Collier, C.; Hay, P. The roles and development needs of teachers to promote creativity: A systematic review of literature. *Teach. Teach. Educ.* **2014**, *41*, 34–41. [[CrossRef](#)]
48. Lo, H.K.; McCord, M.R. Routing through dynamic ocean currents: General heuristics and empirical results in the gulf stream region. *Transp. Res. Part B Methodol.* **1995**, *29*, 109–124. [[CrossRef](#)]
49. Lu, R.; Turan, O.; Boulougouris, E.; Banks, C.; Incecik, A. A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping. *Ocean Eng.* **2015**, *110*, 18–28. [[CrossRef](#)]
50. MAN Diesel. *Turbo. Basic Principles of Ship Propulsion*; MAN Diesel & Turbo: Copenhagen, Denmark, 2011.
51. Kwon, Y.J. The Effect of Weather, Particularly Short Sea Waves, on Ship Speed Performance. Ph.D. Thesis, Newcastle University, Newcastle, UK, 1981.
52. Kuroda, M.; Sugimoto, Y. Evaluation of ship performance in terms of shipping route and weather condition. *Ocean Eng.* **2022**, *254*, 111335. [[CrossRef](#)]
53. Lee, J.; Yoo, S.; Choi, S.; Kim, H.; Hong, C.; Seo, J. Development and application of trim optimization and parametric study using an evaluation system (solution) based on the rans for improvement of EEOI. In Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering, San Francisco, CA, USA, 8–13 June 2014.
54. Moustafa, M.M.; Yehia, W.; Hussein, A.W. Energy efficient operation of bulk carriers by trim optimization. In Proceedings of the International Conference on Ships and Shipping Research, Lecco, Italy, 24–26 June 2015.
55. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
56. Zhu, X.; Chu, J.; Wang, K.; Wu, S.; Yan, W.; Chiam, K. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. *J. Rock Mech. Geotech. Eng.* **2021**, *13*, 1231–1245. [[CrossRef](#)]
57. Budholiya, K.; Shrivastava, S.K.; Sharma, V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 4514–4523. [[CrossRef](#)]
58. Davagdorj, K.; Pham, V.H.; Theera-Umpon, N.; Ryu, K.H. XGBoost-based framework for smoking-induced noncommunicable disease prediction. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6513. [[CrossRef](#)]
59. Li, W.; Yin, Y.; Quan, X.; Zhang, H. Gene expression value prediction based on XGBoost algorithm. *Front. Genet.* **2019**, *10*, 1077. [[CrossRef](#)]
60. Song, K.; Yan, F.; Ding, T.; Gao, L.; Lu, S. A steel property optimization model based on the XGBoost algorithm and improved PSO. *Comput. Mater. Sci.* **2020**, *174*, 109472. [[CrossRef](#)]
61. Zheng, H.; Wu, Y. A XGboost model with weather similarity analysis and feature engineering for short-term wind power forecasting. *Appl. Sci.* **2019**, *9*, 3019. [[CrossRef](#)]
62. Duan, T.; Anand, A.; Ding, D.Y.; Thai, K.K.; Basu, S.; Ng, A.; Schuler, A. Ngboost: Natural gradient boosting for probabilistic prediction. In Proceedings of the International Conference on Machine Learning, Online, 13–18 July 2020.
63. Zhao, J.; Shi, M.; Hu, G.; Song, X.; Zhang, C.; Tao, D.; Wu, W. A data-driven framework for tunnel geological-type prediction based on TBM operating data. *IEEE Access* **2019**, *7*, 66703–66713. [[CrossRef](#)]
64. Karagiannidis, P.; Themelis, N. Data-driven modelling of ship propulsion and the effect of data pre-processing on the prediction of ship fuel consumption and speed loss. *Ocean Eng.* **2021**, *222*, 108616. [[CrossRef](#)]
65. Akyuz, E.; Cicek, K.; Celik, M. A comparative research of machine learning impact to future of maritime transportation. *Procedia Comput. Sci.* **2019**, *158*, 275–280. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.