

Article

AI-Driven Cost Estimation in Shipbuilding: A Hybrid Framework Integrating PSO and XGBoost

Yifan Su ¹, Guanghua Xu ^{1,*} and Jing Yang ²

¹ School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China; syf940126@163.com

² Finance Department, Zhongchuan Chengxi Shipyard Co., Ltd., Jiangyin 225217, China; 231210401120@stu.just.edu.cn

* Correspondence: njust17315294386@163.com

Abstract

Accurate shipbuilding cost estimation is important for project quotation, budget preparation, cost control, and operational decision-making. This study proposes an AI-driven hybrid framework for shipbuilding cost estimation by integrating particle swarm optimization (PSO) with the extreme gradient boosting (XGBoost) algorithm. The empirical dataset consists of 51 conventional vessels constructed by a shipyard between 2020 and 2024. Based on the characteristics of these vessels and their construction environment, 11 internal and external cost-influencing factors were selected as input variables. The dataset was divided into a training set and a test set at a ratio of 85% to 15%. The PSO-XGBoost model was compared with the conventional XGBoost model using R^2 and MAPE. On the test set, the proposed model achieved an R^2 of 0.8718 and a MAPE of 3.87%, compared with 0.8474 and 4.15%, respectively, for the conventional XGBoost model. SHAP analysis was further employed to examine the direction and relative contribution of the influencing factors, thereby improving the interpretability of the estimation results. Case-based validation using different conventional ship types further demonstrates the practical feasibility of the proposed framework. The findings indicate that PSO-based hyperparameter optimization can improve the predictive performance of XGBoost and provide interpretable decision support for shipbuilding cost management.

Keywords: shipbuilding cost estimation; PSO-XGBoost; SHAP analysis; interpretable machine learning; cost management

1. Introduction

The shipbuilding industry occupies a strategic position in international trade and the marine economy, serving as a core pillar that connects the global supply chain. Its level of development directly reflects a country's industrial strength and economic influence. Prompted by global economic recovery and the implementation of environmental protection regulations, the demand for international shipping has significantly increased, leading to a surge in new ship orders. This trend has garnered widespread attention regarding the future direction of the shipbuilding industry. Currently, the three major indicators of the global shipbuilding sector are showing an upward trajectory. In 2024, the total shipbuilding completion volume reached approximately 81.86 million deadweight tons and 38.4 million corrected gross tons, representing a year-on-year growth of 3.4% in corrected gross tons. The newly received order volume was approximately 152.13 million deadweight tons and 60.88 million corrected gross tons, reflecting increases of 19.6%



Academic Editor: Lars Johanning

Received: 6 July 2026

Revised: 20 August 2026

Accepted: 28 August 2026

Published: 31 August 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and 22.4%, respectively, year-on-year. By the end of the year, the order volume on hand was approximately 288.17 million deadweight tons and 135.2 million corrected gross tons, showing year-on-year increases of 14% and 8.5%, respectively.

However, prolonged ship delivery cycles and fluctuations in raw material prices have exerted significant pressure on profit margins, leaving the overall profitability of the shipbuilding industry at a relatively low level of 5% to 10%. Shipbuilding enterprises face multiple challenges in cost control, including rising steel prices, increased investment in green technologies, and the demands of digital transformation and upgrading. At the same time, the global shipbuilding market remains intensely competitive, with traditional shipbuilding powerhouses (such as Japan and South Korea) and emerging players (such as China and Vietnam) competing across multiple dimensions, including technological innovation and cost efficiency. For shipbuilding enterprises, accurate cost estimation plays a critical role in clearly identifying all construction-related expenses, determining product pricing, and ensuring profitability. Therefore, to sustain the competitiveness and market position of the domestic shipbuilding industry in the highly contested market, it is essential to develop more precise and reasonable cost estimation methods for ship construction.

The current shipbuilding market faces severe challenges, including stringent safety regulations, intense market fluctuations, demanding energy-saving requirements, and increasingly rigorous environmental protection standards. At the same time, advances in marine engineering increasingly rely on sophisticated numerical analysis and performance optimization to address complex hydrodynamic, aerodynamic, and fluid–structure interaction problems [1–3]. Such technological development is also accompanied by growing attention to economic viability and cost management, as highlighted in recent studies of emerging marine energy technologies [4]. Against this broader background, the empirical reference value of traditional cost estimation methods has significantly diminished, making cost estimation increasingly complex and difficult [5].

Traditional methods primarily rely on simple linear formulas and proportional techniques. For example, a major Chinese shipyard uses a formula where Fixed Costs = Manufacturing expenses $\times$ 88% + Employee compensation $\times$ 100% + Kinetic energy fees $\times$ 15% + Slipway fees $\times$ 100% + Terminal fees $\times$ 95% + Tugboat fees $\times$ 80.41% + Dry dock fees $\times$ 95% + Large floating crane fees $\times$ 100%. Such methods heavily depend on subjective judgment and static historical data, resulting in low accuracy and a lack of dynamic response to market fluctuations. This often leads to a significant disconnect between estimated and actual costs, failing to meet shipowners' personalized needs. Therefore, shipbuilders urgently require more accurate, efficient, and dynamic cost estimation methods. This necessity drives the exploration of AI-driven approaches, such as the hybrid PSO-XGBoost model proposed in this study, to address the complexities of the modern shipbuilding environment [6].

At the same time, the machine learning landscape continues to evolve beyond established tree-based ensemble methods and conventional neural networks. Kolmogorov–Arnold Networks (KANs), inspired by the Kolmogorov–Arnold representation theorem, replace the fixed node-based activation functions and scalar weights of conventional multi-layer perceptrons with learnable univariate functions defined on network edges, offering a promising architecture for improving model accuracy and interpretability [7]. A recent systematic review further demonstrates the rapid development of KAN-related research in terms of basis-function design, accuracy, computational efficiency, regularization, and convergence, while also identifying several unresolved methodological and practical challenges [8]. Within this evolving methodological landscape, PSO-XGBoost is adopted in the present study because it provides a mature and practical framework for structured tabular data, enables systematic hyperparameter optimization, and supports SHAP-based interpre-

tation. Future research may compare PSO-XGBoost with KAN-based models using larger and more diverse shipbuilding datasets to further examine their respective trade-offs in estimation accuracy, interpretability, computational efficiency, and generalization capability.

2. Literature Review

2.1. Research on Enterprise Cost Management

Cost management is a critical component of enterprise management, encompassing a series of scientific management practices such as cost accounting, cost prediction, cost control, and cost management evaluation throughout the production and operational processes of an enterprise [9]. The academic community has extensively explored various aspects of enterprise cost management, focusing primarily on cost prediction, cost control, and the development of diverse cost management methods [10].

In the field of cost prediction, Bodendorf et al. (2022) [11] evaluated the ability of Artificial Neural Networks (ANN) to extract cost structure data through case studies in the automotive industry. A comparative analysis was also conducted between ANN and other promising Machine Learning (ML) algorithms in terms of purchase cost prediction. The findings revealed that certain ML algorithms outperformed ANN in accuracy, delivering superior results for cost prediction and cost structure analysis. Cheng et al. (2021) [12] developed a price prediction model for construction materials using the Particle Swarm Optimization (PSO) algorithm integrated with the Least Squares Support Vector Machine (LSSVM). By leveraging LSSVM to determine the optimal parameter combinations, their model effectively established a pre-cost management framework, enhancing the overall efficiency of cost management. Kim et al. (2021) [13] addressed the issue of pre-event cost forecasting by implementing early warning systems to identify potential cost overruns. They proposed the Multivariate Time Series (MTS) model to achieve accurate cost prediction results.

In the area of cost control, Yang and Tao (2023) [14] tackled cost control challenges in the logistics transportation process by applying a fast non-dominated sorting genetic algorithm within multi-objective genetic algorithms. Their approach established a logistics route optimization model, improving cost control efficiency while balancing logistics speed and customer satisfaction through sensitivity analysis to achieve an optimal overall objective. Liu et al. (2023) [15] conducted an in-depth study on enterprise cost control using blockchain technology. By focusing on three critical factors—quality, schedule, and cost—they introduced the concept of target cost optimization control and outlined specific control strategies. Zhong et al. (2025) [16] proposed a financial information system architecture based on cloud computing, leveraging its service functions and models. Their method demonstrated the potential to reduce enterprise costs by as much as 60%, while simultaneously enhancing financial management efficiency and supporting decision-making processes related to market operation and enterprise development. Walter (2023) [17] examined an AI-enabled architecture for manufacturing supply chains that integrates artificial intelligence, data analytics, and edge–cloud technologies. The proposed framework supports production monitoring and planning, thereby facilitating more adaptive and efficient supply chain management. Cremer (2020) [18] emphasized the importance of data timeliness in cost control processes. He highlighted the need for modern information management systems during project implementation to ensure accurate and timely data collection, thereby minimizing discrepancies between projected and actual data. Li and Ni (2025) [19] introduced an innovative approach that employs the continuous review list model to integrate price information from both spot and futures markets, enabling effective inventory cost management.

In the context of cost management methods, Tambunan (2024) [20] applied the concept of quality cost within the strategic cost management framework, linking quality with cost. This approach provided managers with a tool to assess their departments' strategic contributions to quality cost and improve performance both efficiently and effectively. Rounaghi et al. (2021) [21] developed a strategic cost management model aimed at reducing cost stickiness and enhancing enterprise sustainability. This model introduced market-based pricing strategies to improve profitability and competitiveness for manufacturing companies operating in highly competitive global markets. Pavlatos and Kostakis (2022) [22] explored how innovation interacts with management control using the target costing method. Through a study of 108 large industrial enterprises with R&D departments, they revealed that information systems indirectly influence product innovation via target costing. Rymarzak and Marmot (2020) [23] highlighted the necessity of addressing deviations between dynamic costs and target costs in business operations. They advocated for monitoring and adjusting costs during implementation phases to maximize the achievement of organizational objectives. Zhang and Li (2021) [24] discussed the advantages of Activity-Based Costing (ABC) in project budgeting and managing value chains. Their findings demonstrated that the ABC method extends cost management across an enterprise's entire value chain, overcoming the limitations of traditional costing methods in specific manufacturing processes. Ozcalici et al. (2024) [25] implemented the ABC system to calculate unit cost in the hotel industry, using the Full Consistency Method (FUCOM) to allocate expenses to overnight stays. Araújo et al. (2020) [26] validated the effectiveness of ABC in managing indirect costs through practical applications in agricultural enterprises, demonstrating its ability to accurately measure indirect costs. Duarte et al. (2024) [27] highlighted the versatility of the life cycle costing method, which, beyond its application in early-stage product or system development, also ensures competitiveness and cost efficiency in future military product/system operations.

2.2. Research on Machine Learning Applications

Machine learning has found extensive applications across various fields and has delivered remarkable results. Its areas of application include, but are not limited to, healthcare, finance, construction, engineering, and transportation. At the same time, machine learning is predominantly applied to data-driven tasks, especially in scenarios involving pattern recognition, prediction, classification, or automated decision-making. Among these, its utilization in prediction-related scenarios is particularly diverse and abundant.

In scenarios such as classification and automated decision-making, Lu et al. (2023) [28] proposed a credit risk assessment method for the financial industry based on FISM-K2. This method integrates the Fuzzy Interpretive Structural Model (FISM) for expert knowledge acquisition and the K2 algorithm for big data learning. A Bayesian Network (BN) is employed for both forward and backward reasoning, enhancing the system's predictive capabilities. Similarly, He et al. (2025) [29] explored the nonlinear impact of the chief economist system on the effectiveness of securities companies in serving the capital market. By constructing machine learning models such as decision trees and random forests, the study provides valuable insights into how securities companies can leverage their internal resources to foster capital market development while highlighting the broader economic benefits of the chief economist system to society. Furthermore, Jiang et al. (2025) [30] integrated epidemiological methods with explainable machine learning techniques to investigate the association between urinary organic compound metabolites and the risk of chronic kidney disease. Their analytical approach included multivariate logistic regression, LASSO regression, and five machine learning models: Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), and Multilayer Perceptron

(MLP). To enhance model interpretability, Shapley Additive Explanations (SHAP) analysis was employed, providing deeper insights into the contribution of individual features to the prediction outcomes.

In prediction scenarios, Ting et al. (2025) [31] applied a series of machine learning regression methods to forecast the financial performance of microcredit institutions. These methods included linear regression, partial least squares, stepwise-selected linear regression, elastic net, random forests, quantile random forest, Bayesian ridge regression, k-nearest neighbors, and support vector regression. The results indicated that the random forest model emerged as the most suitable choice, outperforming other models in terms of predictive accuracy. Similarly, Kim et al. (2024) [32] investigated the prediction of the Architectural Billings Index (ABI) and demonstrated that the Long Short-Term Memory (LSTM) network achieved the highest accuracy compared to both machine learning and traditional time series prediction models. This superior performance was attributed to LSTM's ability to address vanishing and exploding gradient problems and to learn long-term dependencies in continuous ABI time series data. Yan et al. (2022) [33] proposed an investment cost estimation model for prefabricated concrete buildings based on the XGBoost machine learning algorithm. The study revealed that, in comparison to traditional machine learning methods such as Support Vector Machine (SVM), Back Propagation Neural Network (BP Neural Network), and Random Forest (RF), XGBoost exhibited superior generalization capabilities and interpretability.

Taken together, these studies indicate that the suitability of a machine learning method depends on the characteristics of the prediction problem and the underlying data. Linear regression provides a transparent relationship between input variables and estimated costs, but its predefined linear structure limits its ability to represent complex nonlinear relationships and interactions among cost drivers. Support vector regression can model nonlinear relationships through kernel functions and is applicable to regression problems with relatively limited samples; however, its performance is sensitive to data scaling, kernel selection, and parameter settings, while its prediction mechanism is relatively difficult to interpret. Random forest can capture nonlinear relationships and reduce prediction variance by integrating multiple independently constructed decision trees, but its bagging mechanism does not progressively correct the residual errors generated by preceding trees.

In contrast, XGBoost constructs decision trees sequentially, enabling subsequent trees to learn from the residual errors of the preceding model, while regularization, feature subsampling, and shrinkage help control model complexity. These characteristics are particularly relevant to shipbuilding cost estimation, in which vessel dimensions, construction conditions, market indices, exchange rates, and digitalization-related factors may jointly influence total cost through nonlinear relationships and interactions. Moreover, XGBoost can be integrated with SHAP analysis to explain the direction and relative contribution of individual factors. Therefore, considering both the findings of previous comparative studies and the characteristics of shipbuilding cost data, XGBoost was selected as the basic estimation model in this study, and PSO was further introduced to systematically optimize its key hyperparameters. The theoretical comparison is summarized in Table 1.

In cost prediction scenarios, Huang et al. (2024) [34] innovatively developed the Adaptive Self-Explanatory Convolutional Neural Network (ASCNN) to enhance the accuracy and efficiency of specific cost estimation for rural residences. The ASCNN model integrates a random forest algorithm to identify key features and employs PSO to fine-tune parameters, thereby improving predictive performance. To ensure credibility and transparency, the model interprets its decision-making process through SHAP analysis. Similarly, Paik et al. (2025) [35] proposed a cost prediction model tailored for the early-stage planning of pavement maintenance projects. The study utilized the XGBoost algorithm, random for-

est model, and tree regression algorithm to construct the prediction framework. The results demonstrated that the additional tree regression model, combined with SelectFromModel for feature selection, achieved superior predictive performance. Zhang et al. (2025) [36] introduced a data-driven framework designed to provide conceptual cost estimates for infrastructure projects. The framework integrates feature selection, XGBoost, Bayesian Optimization (BO), and SHAP analysis to not only enhance cost prediction performance but also improve result interpretability for early decision-making. Deepa et al. (2025) [37] proposed a hybrid model that combines machine learning techniques with ANN, achieving a 97.42% accuracy rate in predicting the cost of pile foundations. Hammann (2024) [38] explored the applicability of machine learning and big data technologies in cost estimation for complex products, such as passenger vehicles in the automotive manufacturing industry. The study employed two advanced machine learning algorithms, namely the artificial neural network (ANN) and the Gradient Boosting Regression (GBR) model, to estimate and analyze direct material cost, offering insights into cost optimization for complex manufacturing processes. Srinadhraju et al. (2024) [39] optimized the Constructive Cost Model (COCOMO) by integrating the Social Group Optimization (SGO) algorithm. The experimental results demonstrated that the SGO-optimized COCOMO significantly improved the accuracy of software cost predictions. Similarly, Draz et al. (2024) [40] developed a software cost prediction model that combines convolutional neural networks (CNN) with PSO in the context of time series prediction. This model enables automatic feature extraction and hyperparameter adjustment, thereby reducing the need for manual parameter selection and facilitating fine-tuning for improved accuracy.

Table 1. Theoretical comparison of commonly used regression models.

Model	Modeling Characteristics	Main Advantages	Limitations and Relevance to This Study
Linear regression (LR)	Estimates an additive linear relationship between predictors and the response.	Transparent, computationally efficient, and easy to interpret.	Limited in representing nonlinear relationships and interactions among shipbuilding cost drivers; useful mainly as a simple baseline.
Support vector regression (SVR)	Uses kernel functions and a maximum-margin principle for nonlinear regression.	Effective for nonlinear problems and relatively limited samples.	Sensitive to data scaling, kernel choice, and parameter settings, while its prediction mechanism is comparatively difficult to interpret.
Random forest (RF)	Combines independently constructed randomized decision trees through bagging.	Captures nonlinear relationships, reduces variance, and is relatively robust to noise.	Its independent trees do not sequentially correct preceding residual errors, and the ensemble may provide less targeted error reduction.
XGBoost	Builds gradient-boosted trees sequentially with regularization, shrinkage, and feature subsampling.	Models nonlinear interactions, progressively corrects residual errors, controls complexity, and supports SHAP interpretation.	Performance depends on hyperparameter selection; therefore, it is adopted as the base model and further optimized using PSO in this study.

Source(s): Authors' synthesis based on Ting et al. (2025) [31] and Yan et al. (2022) [33].

2.3. Research on Shipbuilding Cost Estimation

Compared with studies on enterprise cost management and the application of machine learning, research specifically addressing shipbuilding costs remains relatively limited in the academic community. Fafandjel et al. (2010) [41] proposed a computer-integrated

cost structure optimization model aimed at defining the optimal and most profitable cost structure for ship production, thereby enabling improved cost control within the model's parameters. Jiang and Strandenes (2012) [42] examined the sources and constraints of China's cost advantage in shipbuilding, evaluating changes in shipbuilding costs and the competitiveness of China's shipbuilding industry over time. Broderick et al. (2013) [43] explored methods to minimize coating costs by examining the shipbuilding process and focusing on paint application during the shipbuilding stage. Carberry (2024) [44] highlighted that 80% of the actual cost increases in shipbuilding budgets can be attributed to rising unit costs, while the remaining 20% stem from changes in composition. Tran et al. (2025) [45] introduced a novel model leveraging ship navigation data to quantify the impact of container fleet stranding on overall shipbuilding costs.

Meanwhile, compared with the extensive application of various cost estimation methods in fields such as software engineering, construction engineering, and manufacturing, research on their application in the shipbuilding industry remains relatively limited, with methodologies often lacking diversity. Alblas and Pruyn (2024) [46] introduced a concept leveraging graph theory to integrate process information into cost estimation for shipbuilding steel structures. This approach is particularly suited to the early stages of design projects that are based on orders. Gualeni and Vairo (2023) [47] applied the Bayesian method within the framework of the General Linear Model (GLM), using individual items or components of a larger system, to establish a quantitative relationship between increased maintenance costs and reduced engine room clearance in ships. This approach enables the estimation of maintenance costs during the ship's engine room design process. Li et al. (2019) [48] proposed a novel three-stage method for estimating shipbuilding construction times, utilizing data mining techniques such as the K-Means algorithm, Data Envelopment Analysis (DEA), and a genetic algorithm-backpropagation neural network (GA-BPNN) model. Lin and Shaw (2017) [49] developed an innovative feature-based cost estimation method, which calculates shipbuilding cost based on preliminary specifications, including steel, other major materials, engines, generators, core equipment, and labor hours. Azhar and Kristiyono (2022) [50] employed linear regression methods to process ship data and estimate the construction costs of ferries and their components.

In summary, existing research on enterprise cost management has primarily focused on in-process cost control and the application of different cost management methods, while comparatively less attention has been given to cost estimation. Although machine learning has been widely applied to cost estimation in fields such as software engineering, complex-product manufacturing, and construction, its application in the shipbuilding industry remains relatively limited. Existing shipbuilding-related studies generally focus on individual cost components, ship prices, maintenance costs, or specific stages of ship construction, with relatively limited consideration of total shipbuilding cost. Moreover, the combined effects of internal vessel and project characteristics and external market conditions, as well as the interpretability of prediction results, have not been sufficiently examined. These issues indicate the need for a more comprehensive and interpretable data-driven framework for total shipbuilding cost estimation.

2.4. Novelty and Contributions

Although PSO-XGBoost has been applied in other research fields, the contribution of the present study lies in adapting and integrating this approach according to the specific characteristics and practical requirements of total shipbuilding cost estimation. The contributions of this study are mainly reflected in the following three aspects.

First, this study formulates total shipbuilding cost estimation as a data-driven prediction problem based on multiple indirect but practically obtainable cost drivers. The

proposed indicator system jointly incorporates 11 factors covering vessel characteristics, construction conditions, external market fluctuations, exchange rates, and the degree of digitalization. This allows shipbuilding cost to be estimated without relying exclusively on detailed direct cost components, such as actual labor and major equipment costs, which may not be fully determined during the early estimation stage.

Second, this study develops a PSO-XGBoost-based shipbuilding cost estimation framework according to the nonlinear and heterogeneous characteristics of shipbuilding cost data. XGBoost is employed to identify the complex nonlinear relationships and interactions among multiple cost-influencing factors, while PSO is used to jointly optimize the key hyperparameters of XGBoost. This integration reduces the dependence of parameter selection on repeated manual trial and error and provides a systematic method for configuring the prediction model in the context of shipbuilding cost estimation.

Third, SHAP analysis is integrated into the estimation framework to quantify the direction and relative contribution of each influencing factor. This makes it possible to identify the principal drivers of shipbuilding cost and to explain the model's prediction results. Consequently, the proposed framework extends beyond cost prediction by connecting parameter optimization and total-cost estimation with cost-driver interpretation and management decision support.

Accordingly, the novelty of this study arises from the domain-specific formulation and systematic integration of the analytical framework, rather than from modifying the underlying PSO or XGBoost algorithms. The framework combines multi-source cost drivers, automated hyperparameter optimization, total-cost prediction, and interpretable analysis in a manner tailored to the characteristics of shipbuilding cost management.

3. Methodology

3.1. Principles of the XGBoost Algorithm

XGBoost is a sophisticated ensemble learning algorithm, with its core mechanism rooted in the gradient boosting framework. This algorithm constructs a series of decision trees iteratively, where the final output is obtained as the cumulative sum of the predicted values across all decision trees. The decision trees in the XGBoost algorithm typically follow the CART (Classification and Regression Tree) structure, which consists of internal nodes, branches, and leaf nodes. Each internal node represents a splitting condition based on a specific feature. Depending on the value of the sample for that feature, the sample is directed to either the left subtree or the right subtree. Ultimately, samples are assigned to a leaf node, which outputs the corresponding estimated values, such as shipbuilding costs [51]. The specific structure of the decision tree in the XGBoost algorithm is illustrated in Figure 1.

The core idea of XGBoost is rooted in the concept of “boosting”, where a strong learner is constructed by combining multiple weak learners using an additive training approach [52]. The mathematical expression of the shipbuilding cost estimation model based on XGBoost is presented in Formula (1):

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

Here, x_i represents the sample of the ship data, $\hat{y}_i$ denotes the predicted values of the shipbuilding costs, K is the total number of the current tree, F represents the space of all possible tree models, and f_k refers to a specific tree model within K . In each iteration, a new tree model is added to the original model without affecting its existing structure, and

the new model is combined with the residuals of the original predictions to minimize the residual errors. The iterative process is defined in Formula (2):

$$\begin{cases} \hat{y}_i^0 = 0 \\ \hat{y}_i^1 = f_1(x_i) = \hat{y}_i^0 + f_1(x_i) \\ \vdots \\ \hat{y}_i^k = f_1(x_i) = \hat{y}_i^{k-1} + f_t(x_i) \end{cases} \quad (2)$$

The objective functions of the XGBoost-based shipbuilding cost estimation model are defined in Formulas (3) and (4):

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

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T w_j^2 \quad (4)$$

Among these, $\sum_{i=1}^n l(\hat{y}_i, y_i)$ represents the loss function of the shipbuilding cost estimation model, $\sum_{k=1}^K \Omega(f_k)$ denotes the regularization term of the cost estimation model, T is the number of leaf nodes in the cost estimation model tree, and w is the value of a leaf node. Additionally, γ and λ are the weight coefficients used to control the number of leaf nodes and the values of the leaf nodes, respectively. To accelerate the convergence of the objective function, the loss function is approximately replaced by its second-order Taylor expansion. After rearranging the terms, the objective function at the t -th iteration is presented in Formula (5):

$$L_{obj}^t = \sum_{i=1}^n [l(\hat{y}_i^{t-1}, y_i) + g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i)] + \Omega(f_t) \quad (5)$$

Among these, $g_i = \partial_{\hat{y}_i^{t-1}} l(y_i, \hat{y}_i^{t-1})$, $h_i = \partial_{\hat{y}_i^{t-1}}^2 l(y_i, \hat{y}_i^{t-1})$. When the model is determined after $t - 1$ iterations, the objective function, with the constant term removed, is presented in Formula (6):

$$\tilde{L}_{obj}^t = \sum_{i=1}^n [g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i)] + \Omega(f_t) \quad (6)$$

Define the set $I_j = \{i | q(x_i) = j\}$ as all the values I_j classified into the J -th leaf node. The set is then extended according to the regularization term, as shown in Formulas (7) and (8):

$$\tilde{L}_{obj}^t = \sum_{i=1}^n [g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i)] + \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T w_j^2 \quad (7)$$

$$= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (8)$$

Among these, when calculating the optimal value of the j -th leaf node, as shown in Formula (9):

$$w_j^* = \left(\sum_{i \in I_j} g_i \right) / \left(\sum_{i \in I_j} h_i + \lambda \right) \quad (9)$$

The optimal value of the objective function in the XGBoost-based shipbuilding cost estimation model can then be obtained, as shown in Formula (10):

$$\tilde{L}_{obj}^t = (\sum_{i \in I_j} g_i)^2 / (\sum_{i \in I_j} h_i + \lambda) + \gamma T \quad (10)$$

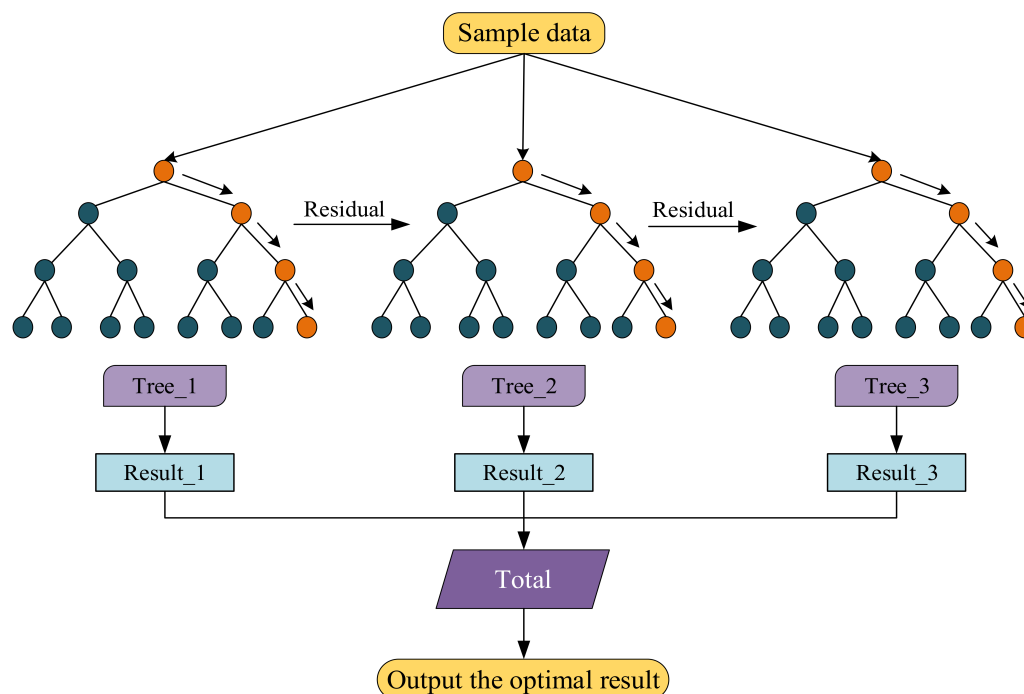


Figure 1. Schematic diagram of the decision tree in the XGBoost algorithm.

3.2. Principles of the PSO Algorithm

Particle Swarm Optimization (PSO) is an optimization algorithm based on swarm intelligence. It simulates the social behavior of a flock of birds or a school of fish, searching for the optimal solution through collaboration and information sharing among individuals (particles). The core concept of PSO lies in adjusting the movement direction of particles in the search space based on their own experience and the collective experience of the group, thereby gradually converging towards the global optimal solution [53]. As illustrated in Figure 2, the optimization process in PSO integrates individual experience with collective intelligence. Specifically, during the search process, each particle considers two key parameters: its own historical best position (P_best) and the global best position (G_best). These two parameters represent the concepts of individual learning and population learning. The PSO algorithm continuously updates the velocity and position of particles using Formulas (11) and (12), enabling the particles to navigate the solution space and iteratively approach the optimal solution for the shipbuilding cost estimation model [54]. To evaluate the quality of each position (i.e., each optimal solution), the algorithm introduces a key metric known as the fitness function. The fitness function directly corresponds to the objective function of the shipbuilding cost estimation model and provides directional guidance for particle movement. This mechanism ensures an effective balance between individual exploration and group cooperation, allowing the particles to efficiently search the solution space and progressively converge towards the optimal solution of the shipbuilding cost estimation model.

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_1 [P_best_{ij}(t) - x_{ij}(t)] + c_2 r_2 [G_best_j(t) - x_{ij}(t)] \quad (11)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (12)$$

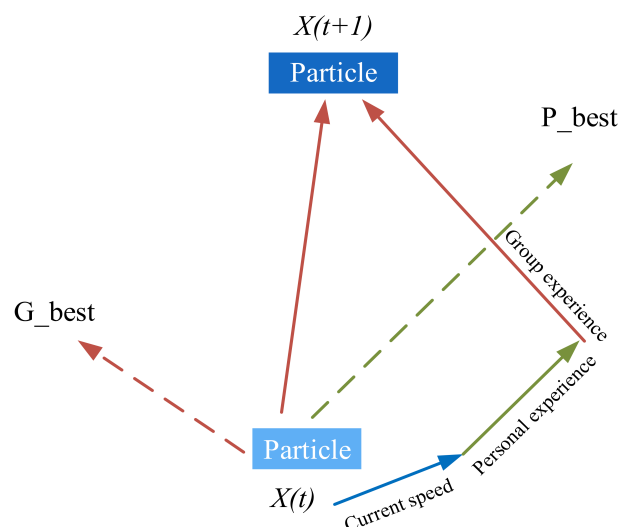


Figure 2. Schematic representation of the PSO algorithm.

Among these, j represents the dimension of the particle, i denotes the i -th particle, and t refers to the current iteration number. The parameters c_1 and c_2 are acceleration constants, typically assigned values within the interval $[0, 2]$, while r_1 and r_2 are two random numbers uniformly distributed in the range $[0, 1]$. As shown in Formulas (11) and (12), c_1 and c_2 integrate the individual-oriented learning and group-oriented learning of particles, enabling particles to simultaneously leverage the search experience of both the individual and the group.

3.3. Development of a Shipbuilding Cost Estimation Model Using the PSO-XGBoost Algorithm

In constructing a model to estimate shipbuilding cost based on the XGBoost algorithm, the accurate selection of parameters plays a critical role in ensuring the reliability of the estimation results. The primary parameters influencing the performance of the XGBoost model include `n_estimators`, `learning_rate`, `max_depth`, `subsample`, and `colsample_bytree`. Detailed descriptions of these parameters are provided in Table 2.

Table 2. Calculation parameters of the XGBoost model.

Parameter	Parameter Description
<code>n_estimators</code>	It indicates the number of trees. This parameter determines the estimation ability of the model. The larger the value is, the stronger the learning and data analysis ability of the model, and the more accurate the estimation results are, but the parameter should not be too large.
<code>learning_rate</code>	It indicates the learning rate. This parameter determines the step size of the iteration. If the parameter is too large, it will affect the accuracy of the operation; if it is too small, it will affect the speed of the operation.
<code>max_depth</code>	It indicates the maximum depth of the tree. This parameter is used to control the overfitting of the algorithm. The larger its value is, the more specific the learning of the model will be, but it should not be too large.
<code>subsample</code>	This parameter is used to control the proportion of random sampling for each tree.
<code>colsample_bytree</code>	This parameter is used to control the proportion of columns randomly sampled for each tree.

In the process of shipbuilding cost estimation using the traditional XGBoost model, higher fitting accuracy is often achieved through manual parameter tuning. However, this approach is inevitably influenced by subjective human factors, which increases the risk of falling into a local optimum. To achieve more accurate and reliable shipbuilding cost estimation results, this study integrates the optimization principles of the PSO algorithm with the XGBoost model. Specifically, the PSO algorithm is employed to optimize the parameters of the XGBoost model, resulting in the development of a PSO-XGBoost-based shipbuilding cost estimation model.

The shipbuilding cost estimation process based on the PSO-XGBoost model is illustrated in Figure 3. This process is highly interconnected and consists of several closely linked steps, which are detailed as follows:

- (1) Based on the analysis of factors influencing shipbuilding costs, the dataset was imported and partitioned, with 85% of the data designated as the training set and 15% as the test set. On this basis, an XGBoost-based shipbuilding cost estimation model was constructed. The model was subsequently trained using the initial parameters. The test set was separated before model development and remained isolated throughout the model-training and PSO-based hyperparameter-optimization processes. It was not used for model fitting, fitness calculation, hyperparameter selection, or the comparison of candidate parameter combinations.
- (2) The PSO algorithm was employed to optimize the parameters of the shipbuilding cost estimation model based on XGBoost. First, the initial parameters of the PSO algorithm were set, and the shipbuilding cost estimation model based on PSO-XGBoost, along with its fitness optimization function, was constructed. The root mean square error (RMSE) of the estimation model under different optimization iterations as the fitness value of each particle, as expressed in Formula (13):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (13)$$

Among these, y_i represents the actual value, $\hat{y}_i$ denotes the predicted value, and N refers to the number of predicted samples. A smaller RMSE value indicates a better optimization effect of the estimation model. During the PSO process, the RMSE used as the particle-fitness value was calculated exclusively from the training data. The independent test data were not involved in the calculation of the fitness values or the selection of the optimal hyperparameter combination.

- (3) Iterative optimization of the main calculation parameters of the XGBoost model is performed using the PSO algorithm. In each iteration, the algorithm dynamically updates the velocity and position of each particle based on its individual optimal value and the global optimal value. This process guides the particle swarm to search for a better combination of XGBoost parameters, thereby gradually improving the model's estimation performance. Finally, the minimum fitness value obtained during the iterative process is selected as the optimal output value.
- (4) After the optimal hyperparameter combination had been determined, the traditional XGBoost model and the PSO-XGBoost model were evaluated using the independent test set. The test set was used only at this final evaluation stage and was not involved in the preceding model-training or PSO-optimization processes. Model performance was assessed using R^2 and MAPE, and the corresponding results were subsequently reported.

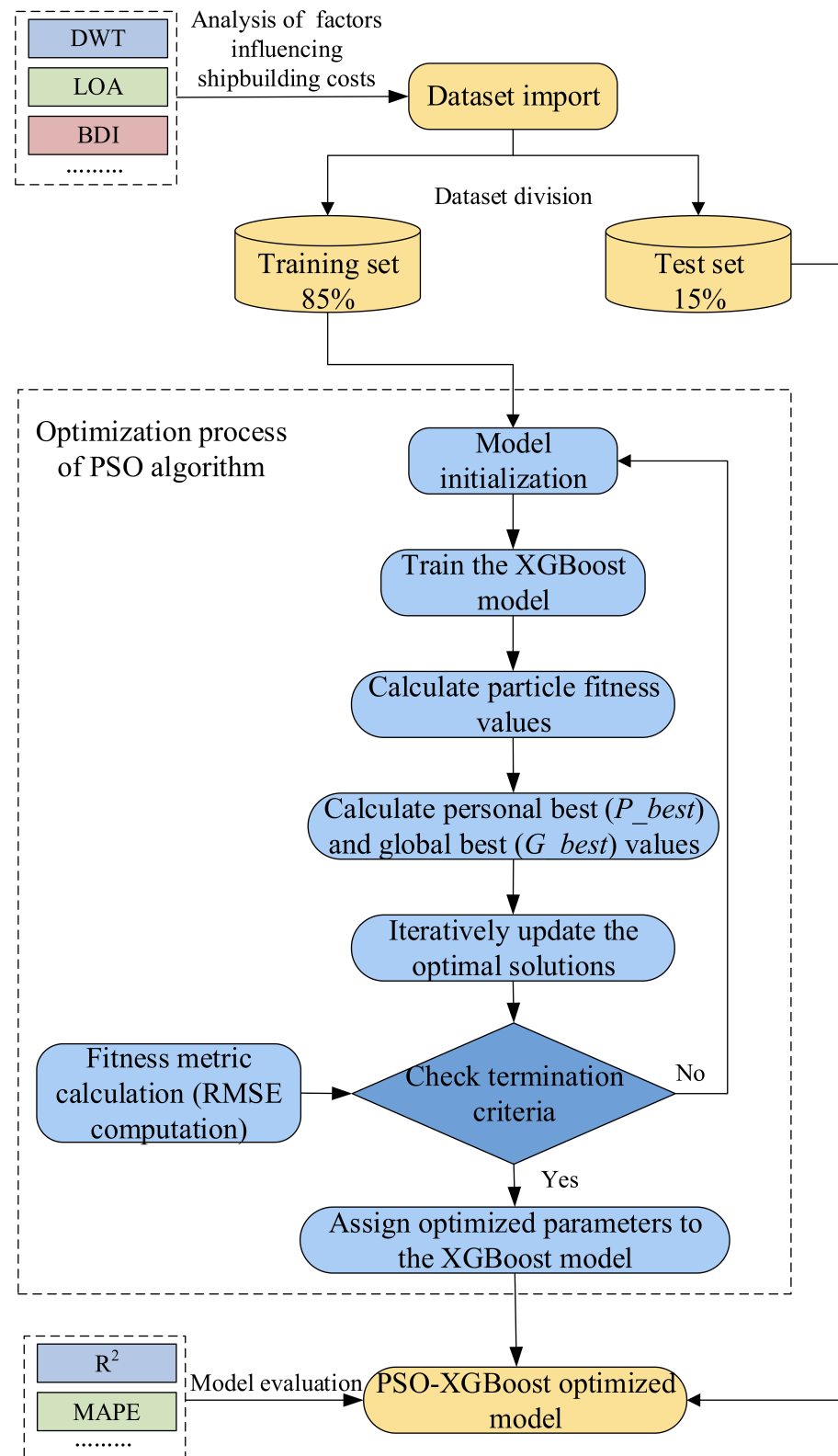


Figure 3. Flowchart of the shipbuilding cost estimation process.

4. Analysis of Influencing Factors

Before identifying the factors used in the estimation model, it is necessary to distinguish between direct cost components and indirect cost drivers. Direct labor cost, material cost, main equipment cost, and manufacturing overhead are accounting components that collectively constitute the total shipbuilding cost. The objective of this study, however, is to

estimate the total cost using upstream technical, production, and market-related factors that are available before these detailed cost components are fully determined. Therefore, the input variables were selected from indirect or ex ante factors rather than from the realized cost components themselves.

Directly using labor, material, or equipment costs as explanatory variables would introduce information derived from the prediction target into the model, potentially resulting in target leakage and circular estimation. Such variables would also be likely to dominate the feature-importance results, making it difficult to identify the underlying technical and business drivers of shipbuilding cost. Accordingly, the exclusion of these direct cost components as model inputs does not imply that their effects are ignored. Instead, the selected upstream variables influence the total shipbuilding cost by affecting material consumption, labor requirements, equipment specifications and procurement, production efficiency, transportation expenses, and manufacturing overhead.

As shown in Figure 4, the cost of shipbuilding is influenced by multiple factors, including production information, ship information, and internal organizational structure [55], as well as external organizational and environmental constraints [56].

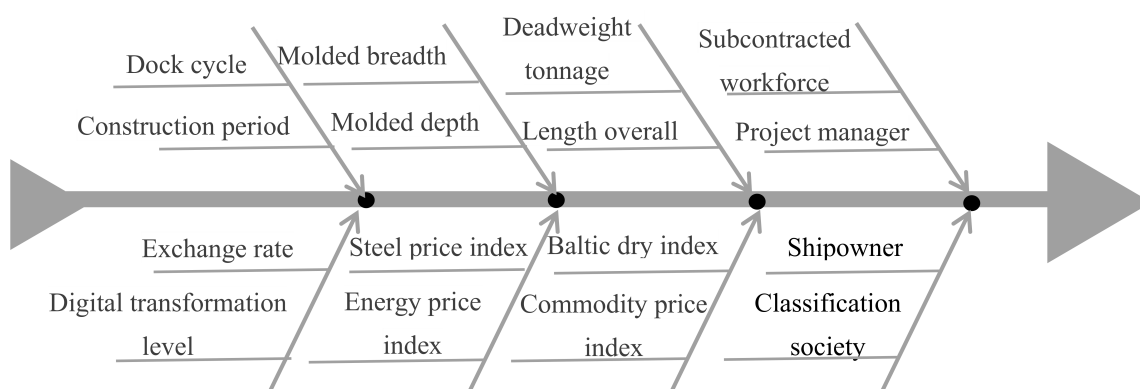


Figure 4. Fishbone diagram of factors Influencing Shipbuilding costs.

To comprehensively investigate these influences, this study adopts a dual-perspective approach, analyzing both internal and external factors. From an internal perspective, it integrates the characteristics of shipbuilding and construction processes, taking into account factors such as ship design and production stages, and identifies key parameters, including deadweight tonnage, length overall, molded breadth, and molded depth. From an external perspective, the study evaluates the impact of external market price fluctuations on shipbuilding costs, focusing on indices such as the Baltic Dry Index (BDI), commodity price index, and steel price index. In addition to these factors, the rapid development and increasing maturity of digital technologies—such as computer technology, artificial intelligence, big data, cloud computing, and the Internet of Things—have significantly expanded the application scenarios of digital transformation. This technological progress has driven profound changes across all aspects of ship design, manufacturing, and operation, reshaping the industry as a whole. Digital transformation has enabled the shipbuilding industry to improve production efficiency, quality control, resource allocation, and other critical aspects, thereby exerting an important influence on cost management in ship construction. Taking this into account, the degree of digital transformation in the shipbuilding industry is also considered one of the key influencing factors in this study. Based on the above analysis, this study identifies 11 major factors affecting shipbuilding costs, which are summarized in Table 3.

Table 3. Factors affecting shipbuilding costs.

Perspective	Influencing Factor	Principal Cost Components Affected	Description
Internal	Deadweight tonnage	Material and equipment costs	Deadweight tonnage affects hull structure, cargo-capacity requirements, material consumption, and the capacity and configuration of relevant equipment, thereby influencing material and equipment costs.
	Length overall	Material, equipment, and labor costs	The inherent volumetric characteristics of the hull, including its length, breadth, and depth, directly influence the cost of raw materials, equipment, and labor, thereby impacting the overall cost of shipbuilding.
	Molded breadth		
	Molded depth		
	Construction period	Labor costs and manufacturing overhead	The construction period affects production scheduling and the duration of labor, facility, and equipment utilization, thereby influencing direct labor costs and manufacturing overhead.
External	Baltic dry index	Material and equipment transportation costs	Fluctuations in the BDI reflect changes in the international shipping market and may affect the transportation costs of shipbuilding materials and equipment.
	Commodity price index	Material costs	Shipbuilding depends heavily on bulk raw materials. Fluctuations in commodity prices therefore affect material procurement costs and, consequently, total shipbuilding cost.
	Steel price index	Material costs	Steel is one of the most extensively used materials in shipbuilding, and fluctuations in steel prices directly affect hull material costs.
	Energy price index	Manufacturing overhead	Cutting, welding, lifting, and equipment operation consume electricity and fuel. Energy-price fluctuations therefore affect manufacturing overhead.
	USD/CNY exchange rate	Material and equipment costs	When imported materials and equipment are settled in US dollars, exchange-rate fluctuations affect their procurement costs when converted into the local currency.
	Degree of digitalization	Labor costs and manufacturing overhead	Digital transformation can streamline design and production processes, improve labor productivity and resource utilization, and reduce rework, thereby influencing labor costs and manufacturing overhead.

5. Result Analysis

5.1. Data Sources

Given that traditional or low-end ship types, such as bulk cargo ships and sawdust boats, have mature designs and relatively complete historical construction data, it is convenient to collect the relevant data required for shipbuilding cost estimation. Consequently, this study selects data from 51 traditional/low-end ships, including bulk cargo ships and sawdust boats, constructed by a specific shipyard between 2020 and 2024, as the research sample. In accordance with data collection requirements, six types of internal data were compiled for these 51 completed ships: deadweight tonnage, length overall, molded breadth, molded depth, construction period, and actual cost. Additionally, six external factors were included: Baltic dry index, commodity price index, steel price index, energy price index, USD/CNY exchange rate, and the degree of digitalization.

Among these factors, the degree of digitalization is difficult to observe directly using a single conventional indicator. Therefore, it was quantified as a numerical text-based proxy using the word-frequency method. Following Xue and Zhang [57], the textual materials used for the measurement consisted of publicly available reports and disclosure documents concerning digital transformation in the shipbuilding industry during the sample period. To ensure comparability across years, the same criteria regarding publication year, industry relevance, and document type were applied when collecting and processing the textual materials.

A fixed keyword dictionary was constructed across four dimensions: digital technology application, Internet-based business models, intelligent manufacturing, and modern information systems. Representative keywords included cloud computing, big data, artificial intelligence, blockchain, the Internet of Things, industrial Internet, e-commerce, intelligent manufacturing, intelligent production, smart factory, intelligent logistics, mod-

ern information systems, data platforms, information sharing, information management, integrated control, and integrated systems. These keywords are also visualized in Figure 5.



Figure 5. Word cloud of High-Frequency terms in digital transformation.

All textual materials were processed using the same procedure. After text extraction and cleaning, a Python 3.14-based program was used to automatically identify and count the predefined digitalization-related keywords. The frequencies of the keywords across the four dimensions were summed to obtain the annual digitalization indicator. All keywords were assigned equal weights, and no manual adjustment was made to the calculated values. Each vessel was assigned the digitalization value corresponding to its construction-start year. Accordingly, vessels starting construction in the same year were assigned the same annual value; the values corresponding to 2020, 2021, 2022, 2023, and 2024 were 252, 268, 258, 163, and 163, respectively. Therefore, this variable represents the digital environment of the shipbuilding industry during the year in which vessel construction commenced rather than a vessel-specific digitalization attribute.

The use of consistent textual sources, a predefined keyword dictionary, standardized text-processing rules, an automated counting procedure, and a uniform temporal-matching method reduces the influence of subjective judgment and makes the measurement transparent and repeatable. Nevertheless, this indicator should be interpreted as a text-based proxy for the degree of digitalization rather than a direct measurement of actual digital investment or implementation effectiveness.

After the aforementioned data collection and processing, the initial input dataset was finalized. Table 4 displays the data for 11 influencing factors and their corresponding actual cost values for 51 ships (due to space constraints, only a portion of the dataset is presented here).

To provide a more comprehensive description of the dataset, Table 5 summarizes the descriptive statistics of all 11 influencing factors based on the 51 vessel samples. The sample size, minimum, maximum, range, mean, and standard deviation are reported to characterize the distribution and variability of each input variable.

Table 4. Partial input dataset.

Ship Type	Deadweight Tonnage (t)	Length Overall (m)	Molded Breadth (m)	Molded Depth (m)	Construction Period (day)	Baltic Dry Index	Commodity Price Index	Steel Price Index	Energy Price Index	USD/CNY Exchange Rate	Degree of Digitalization (Frequency)	Actual Cost
Bulk Carrier	63,500	199.90	32.26	18.90	289	1416.05	172.96	114.89	153.96	6.9854	163	198,243,173.66
Bulk Carrier	85,000	229.90	36.00	20.15	405	2475.05	175.76	154.26	129.28	6.5186	268	246,578,074.82
Bulk Carrier	82,000	229.00	32.26	20.35	289	1760.80	186.68	140.45	156.81	6.3556	258	209,251,844.81
Sawdust Carrier	70,000	215.00	37.00	25.00	295	2943.05	209.51	144.87	204.33	6.6990	258	219,833,416.96
Sawdust Carrier	64,000	210.00	37.00	23.50	414	1410.04	182.32	129.86	162.52	6.8909	163	198,922,435.17
Sawdust Carrier	70,000	215.00	37.00	25.00	298	2464.09	206.86	146.81	200.79	6.3446	258	226,500,655.30
Heavy-Lift Multipurpose Vessel	62,000	199.90	32.26	19.30	329	1650.43	183.15	119.32	166.85	7.1935	163	233,858,847.45
Heavy-Lift Multipurpose Vessel	62,000	199.90	32.26	19.30	383	1831.45	183.21	117.61	166.61	7.2226	163	233,291,876.49
Heavy-Lift Multipurpose Vessel	62,000	199.90	32.26	19.30	392	1081.77	171.35	115.60	149.10	7.1614	163	238,134,189.01
Asphalt Carrier	17,000	160.00	23.50	13.20	434	1487.14	192.21	120.92	173.79	7.0195	258	178,773,742.65
Asphalt Carrier	37,000	179.90	30.60	16.80	424	1630.77	142.21	113.09	92.03	6.7254	252	255,716,299.19
Asphalt Carrier	37,000	179.90	30.60	16.80	480	658.35	184.48	126.65	167.12	6.8380	163	234,764,236.56
Product Tanker	50,000	182.70	32.20	19.10	474	1499.60	173.41	134.03	124.85	6.4601	268	243,043,416.21
Refrigerated Container Ship	28,750	195.00	32.20	17.00	56	601.09	126.57	103.47	83.53	7.0205	252	210,654,497.09

Source(s): Authors' own work.

Table 5. Descriptive statistics of the 11 influencing factors.

Influencing Factor	Unit	N	Minimum	Maximum	Range	Mean	Standard Deviation
Deadweight tonnage	t	51	17,000	88,800	71,800	71,022.55	15,910.72
Length overall	m	51	160.00	229.90	69.90	214.67	17.14
Molded breadth	m	51	23.50	37.00	13.50	33.82	2.69
Molded depth	m	51	13.20	25.00	11.80	20.39	2.26
Construction period	day	51	56	813	757	372.57	104.86
Baltic Dry Index	index points	51	601.09	4287.77	3686.68	1872.90	822.80
Commodity price index	index points	51	126.57	209.51	82.94	183.28	16.77
Steel price index	index points	51	103.47	163.63	60.16	128.22	14.78
Energy price index	index points	51	83.53	204.33	120.80	160.65	28.59
USD/CNY exchange rate	CNY/USD	51	6.3446	7.3071	0.9625	6.8468	0.3295
Degree of digitalization	frequency	51	163	268	105	222.24	48.25

Source(s): Authors' own work.

5.2. Optimization Results of Model Parameters

This study utilizes the PSO algorithm to optimize the parameters of the XGBoost-based shipbuilding cost estimation model. By performing iterative optimization, the optimal parameter combination is identified. The root mean square error (RMSE) of the shipbuilding cost estimation model, evaluated under different numbers of optimization iterations, is adopted as the fitness evaluation criterion. The convergence curve illustrating the optimization process is presented in Figure 6.

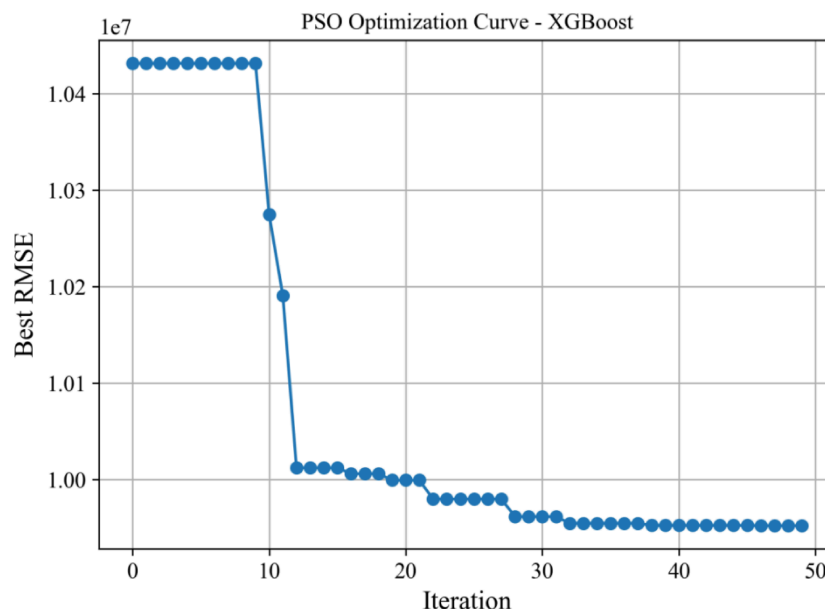


Figure 6. Convergence graph of optimization fitness. Source(s): Authors' own work.

It should be clarified that the fitness values shown in Figure 6 were calculated exclusively from the training data rather than the independent test data. Therefore, neither the PSO convergence process nor the selection of the optimal hyperparameter combination was informed by the test-set performance. The independent test set remained unused until the PSO process and final model construction had been completed.

As shown in Figure 6, during the optimization process of the PSO algorithm, the RMSE value of the shipbuilding cost estimation model gradually decreases with the increasing number of iterations and eventually stabilizes. Specifically, this is reflected in:

During the initial phase of iterations (0–15), the RMSE value of the XGBoost-based shipbuilding cost estimation model dropped significantly, suggesting that the PSO algorithm effectively performed global exploration within the parameter space.

During the middle phase of the iterations (15–30), the rate of decline in the RMSE value of the XGBoost-based shipbuilding cost estimation model gradually slows, indicating that the PSO algorithm transitions into the local fine-tuning search stage.

During the later stage of iterations (30–50), the RMSE value remains essentially stable, suggesting that the PSO algorithm has reached a state of convergence. At this stage, further increasing the number of iterations is unlikely to significantly improve the performance of the XGBoost-based shipbuilding cost estimation model.

The research findings indicate that when the number of iterations reaches 40, the RMSE value of the optimization result approaches its minimum. At this stage, it can be inferred that the PSO algorithm has identified an approximately optimal solution. Finally, the parameter set corresponding to the lowest RMSE value during the iterative process was selected as the optimal parameter combination for the shipbuilding cost estimation model, as presented in Table 6.

Table 6. Optimal parameter combination for the XGBoost-based shipbuilding cost estimation model.

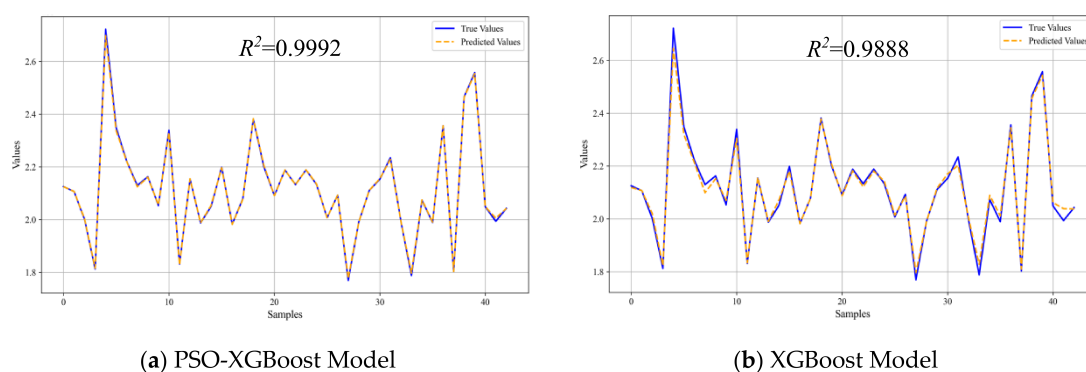
Parameter	Parameter Value
n_estimators	262
learning_rate	0.03
max_depth	9
subsample	0.8005
colsample_bytree	0.8674

Source(s): Authors' own work.

5.3. Comparative Analysis of Model Performance

In this study, a PSO-XGBoost-based shipbuilding cost estimation model was developed using Python programming. The initial ship sample data, which were collected and organized, were imported into Python and processed using the constructed model to estimate shipbuilding cost. Specifically, 85% of the data was designated as the training set and employed to train the estimation model, while the remaining 15% served as the test set to evaluate the model's performance. To further validate the accuracy and stability of the PSO-XGBoost model in estimating shipbuilding costs, this study simultaneously applied the traditional XGBoost model and the PSO-XGBoost model for shipbuilding costs estimation. The estimation results from the two models were then compared and analyzed. Finally, the performance of the models was evaluated using R^2 and MAPE metrics [58].

The R^2 results of the two models for shipbuilding cost estimation are shown in Figures 7 and 8. The closer the data points are to the actual value line, the smaller the discrepancy between the estimated and actual values. R^2 (R-Squared), also known as the coefficient of determination, is a commonly used metric in regression analysis to evaluate the goodness of fit of a model to the observed data. The value of R^2 typically ranges from 0 to 1, with values closer to 1 indicating a better fit, a higher proportion of the variability in the dependent variable explained by the model, and relatively stronger estimation capability. Based on the estimation results of the two models, as shown in Figure 7, the R^2 value of the PSO-XGBoost model on the training set is 0.9992, compared to 0.9888 for the XGBoost model. This demonstrates that the PSO-XGBoost model achieves better performance in shipbuilding cost estimation. An R^2 value of 0.9992 indicates that the PSO-XGBoost model can explain nearly 99% of the variability in the training set, reflecting an excellent fit. As shown in Figure 8, the R^2 value of the PSO-XGBoost model on the test set is 0.8718, compared to 0.8474 for the XGBoost model, further confirming the superior performance of the PSO-XGBoost model. Although the R^2 value on the test set is lower than that on the training set, it still reaches 0.8718, demonstrating that the PSO-XGBoost model's estimates are closer to the actual values, explaining more than 87% of the variability in the test set, and achieving a better fitting effect.

**Figure 7.** R^2 line chart of training sets for the two models. Source(s): Authors' own work.

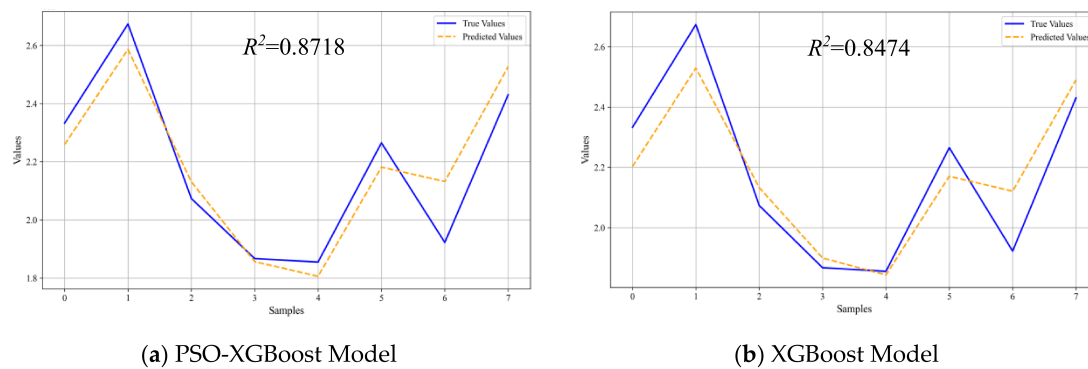


Figure 8. R^2 line chart of test sets for the two models. Source(s): Authors' own work.

To better assess the performance of the models, this study further analyzed the mean absolute percentage error (MAPE) of the shipbuilding cost estimation results generated by the two models. The MAPE results are summarized in Table 7. The PSO-XGBoost model exhibits notably low MAPE values on the training set, indicating that the model error represents an almost negligible proportion of the training data. On the test set, the MAPE value is also as low as 3.87%. In contrast, the traditional XGBoost model achieves an MAPE of 0.68% on the training set and 4.15% on the test set. MAPE is defined as the average percentage of absolute errors between the estimated values and the actual values across all samples. A lower MAPE value reflects higher estimation accuracy, signifying that the model is better at fitting and predicting the actual data. The results of this study demonstrate that the PSO-XGBoost model outperforms the traditional XGBoost model, with significantly lower MAPE values for both the training and test sets. This finding suggests that the PSO-XGBoost model is more adept at capturing the nonlinear characteristics of multiple influencing factors, resulting in superior nonlinear fitting performance for shipbuilding cost estimation. Consequently, the degree of data dispersion is more effectively controlled, leading to more precise and reliable estimation results.

Table 7. MAPE values for the model training and test sets.

Model	Data Set	MAPE Value
XGBoost	Training set	0.68%
XGBoost	Test set	4.15%
PSO-XGBoost	Training set	0.14%
PSO-XGBoost	Test set	3.87%

Source(s): Authors' own work.

A noticeable performance gap can be observed between the training and test sets. The PSO-XGBoost model achieves an R^2 of 0.9992 and a MAPE of 0.14% on the training set, whereas the corresponding values on the test set are 0.8718 and 3.87%, respectively. This difference may be partly attributed to the limited size of the available dataset. Among the 51 vessel samples, 85% were used for model training, and only 15% were reserved for testing. Because the model parameters were learned directly from the training samples, the model achieved a very close fit to the training data. In contrast, the test samples were not involved in model development and may contain combinations of vessel types, principal dimensions, and cost characteristics that are not fully represented in the training set. Moreover, because the test set contains relatively few samples, its evaluation metrics may be sensitive to individual vessels with distinct characteristics.

The observed gap reflects a certain degree of generalization error and suggests that the model may have learned some sample-specific patterns from the training data. Therefore,

the near-perfect training performance should not be interpreted as equivalent predictive performance for unseen samples. Nevertheless, the test-set results, with an R^2 of 0.8718 and a MAPE of 3.87%, indicate that the model retains reasonably good predictive capability for unseen vessel samples within the scope of the current dataset. Accordingly, the test-set performance is considered a more representative measure of the model's practical predictive capability.

Because the test samples were reserved for out-of-sample evaluation and were not used in model training, the test-set results provide an initial check against simple memorization of the training observations. The PSO-XGBoost model achieves an R^2 of 0.8718 and a MAPE of 3.87% on these unseen samples, indicating that the model retains predictive information beyond the training set. Nevertheless, these results do not conclusively rule out overfitting. The substantial decline from the training performance to the test performance, together with the limited number of test samples and the use of a single training–test split, suggests that the evaluation metrics may be sensitive to the specific sample allocation. Therefore, the training-set R^2 of 0.9992 is interpreted only as a measure of in-sample fitting performance rather than evidence of equivalent predictive performance for new vessels. The test-set results are regarded as the more representative indicators of practical predictive capability, and the conclusions are interpreted cautiously within the scope of the current dataset.

It should be emphasized that the comparisons presented in this section are conducted between the conventional XGBoost and PSO-XGBoost models using the same data subsets. Specifically, the two models are compared on the same training set and then separately on the same test set. The comparison is therefore intended to evaluate the improvement achieved through PSO-based hyperparameter optimization rather than to directly compare the training and test metrics of a single model. Given the limited sample size and the use of a single training–test split, the generalizability of the present results should be interpreted with caution.

In terms of computational cost, the PSO-XGBoost model requires additional training time compared with the conventional XGBoost model. The conventional XGBoost model is trained once using a predefined set of hyperparameters, whereas PSO-XGBoost repeatedly trains candidate XGBoost models and evaluates their RMSE values during the iterative hyperparameter-search process. In this study, the PSO procedure was conducted for 50 iterations. As shown in Figure 6, the fitness value became essentially stable after approximately 40 iterations, indicating that further iterations provided only limited improvement while continuing to increase the computational cost.

Nevertheless, the dataset used in this study contains only 51 vessel samples and 11 input factors; therefore, the computational burden associated with each individual XGBoost training process remains relatively limited. More importantly, PSO-based optimization is performed only during the offline model-development stage. Once the optimal parameter combination is obtained, the trained XGBoost model can be directly applied to new cost-estimation tasks without repeating the PSO procedure. The additional training cost is accompanied by an improvement in test-set performance: R^2 increases from 0.8474 to 0.8718, while MAPE decreases from 4.15% to 3.87%. Therefore, although PSO-XGBoost involves a higher offline training cost, the improvement in test-set performance suggests a reasonable trade-off between computational efficiency and estimation accuracy for the shipbuilding cost estimation problem considered in this study.

In summary, the PSO-XGBoost model achieves higher R^2 values and lower MAPE values than the conventional XGBoost model on both the training and test sets. These results indicate that PSO-based hyperparameter optimization improves the fitting and predictive performance of the XGBoost model for shipbuilding cost estimation. However, considering the limited sample size and the observed performance gap between the training

and test sets, the conclusions of this study should be interpreted within the scope of the current dataset.

5.4. SHAP Analysis of Influencing Factors

SHAP analysis is an interpretable machine learning framework grounded in cooperative game theory. Its core principle lies in calculating the contribution of each feature to the model's estimation results, thereby facilitating the understanding of the role each feature plays in decision-making [59]. By computing the Shapley value for each influencing factor, this method can quantitatively assess its contribution to the estimation results of the shipbuilding cost model. Furthermore, SHAP analysis is particularly well-suited for addressing the complex nonlinear relationships inherent in the shipbuilding cost impact system. It not only evaluates the magnitude of each factor's influence on shipbuilding costs but also clarifies its directional impact (positive or negative). This provides valuable insights and decision-making support for shipbuilding enterprises, enabling them to identify key points for cost control and optimize cost management strategies.

This study employed Python programming to perform SHAP analysis on the 11 selected influencing factors mentioned above, as illustrated in Figure 9.

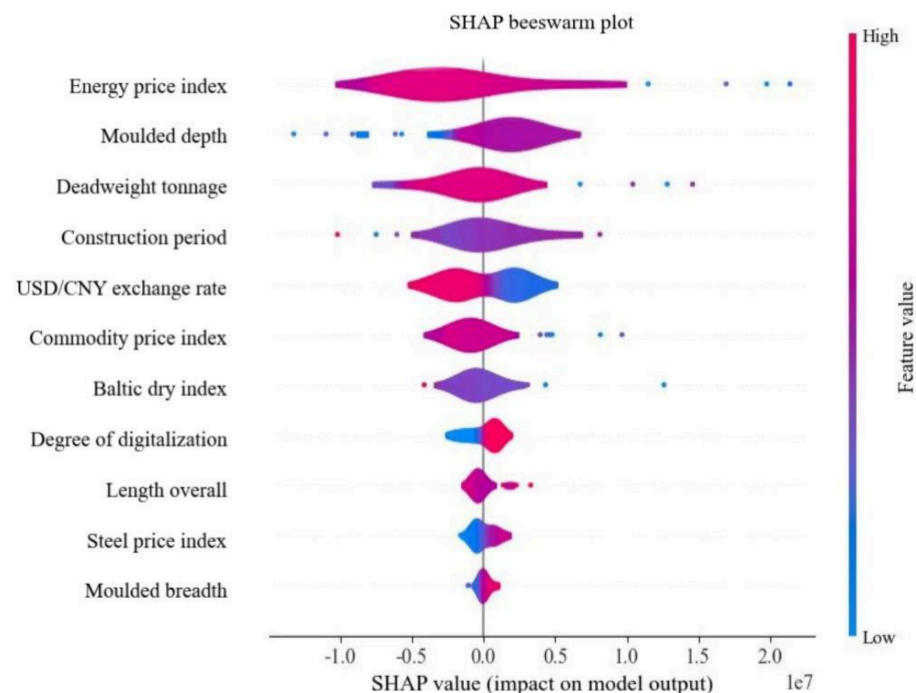


Figure 9. SHAP beeswarm plot of influencing factors in shipbuilding costs. Source(s): Authors' own work.

The vertical axis ranks the factors affecting shipbuilding costs in descending order of importance. The energy price index and molded depth are positioned at the top, indicating that their impact on shipbuilding costs is the most significant. These are followed by deadweight tonnage, construction period, USD/CNY exchange rate, and others. In contrast, length overall, molded breadth, and the steel price index appear at the bottom, suggesting that their impact is relatively minor.

The horizontal axis represents the SHAP value, which reflects both the direction (positive or negative) and the magnitude (absolute value) of the contribution of influencing factors to the output of the cost estimation model. The SHAP values for the energy price index and molded depth exhibit a wide distribution range, with a large number of red dots, indicating that these factors typically have a significant positive impact on shipbuilding cost estimates. Specifically, an increase in the energy price index or molded depth significantly

drives up shipbuilding cost. Similarly, the SHAP values for deadweight tonnage and construction period are widely distributed with many red dots, suggesting that they also exert a substantial positive impact on shipbuilding cost estimates. In contrast, the SHAP values for the exchange rate are widely distributed but exhibit a color gradient from red to blue, which indicates that both increases and decreases in the exchange rate can cause significant positive or negative fluctuations in shipbuilding cost estimates.

The SHAP values for the commodity price index and the Baltic Dry Index are more concentrated and exhibit a narrower distribution range, with a predominance of red dots. This suggests that their increases can significantly push up shipbuilding costs, albeit to a lesser extent than the factors mentioned earlier. The SHAP values for the degree of digitalization show a narrow distribution range with both red and blue dots, indicating that this factor may have a minor positive or negative effect on shipbuilding cost estimates.

In contrast, the SHAP values for length overall, molded breadth, and the steel price index are relatively close to zero, indicating that these variables provide limited additional predictive information within the current model. In particular, the relatively small SHAP value of the steel price index appears counterintuitive, given that steel constitutes a major component of shipbuilding material costs. However, SHAP values represent the incremental predictive contribution of each variable within the current model and dataset after the effects of the other input variables have been considered; they do not directly represent the proportion of a particular cost component in the total shipbuilding cost. Therefore, the small SHAP value should not be interpreted as indicating that steel itself is economically unimportant to shipbuilding.

Several factors may explain this result. First, the steel price index exhibits relatively moderate variation within the current sample. As shown in Table 5, its value ranges from 103.47 to 163.63, with a mean of 128.22 and a standard deviation of 14.78, corresponding to a coefficient of variation of approximately 11.5%. This relatively limited independent variation reduces the amount of cost variation that the model can attribute to this factor. Second, some of the market information represented by the steel price index may overlap with that captured by the commodity price index and other market-related variables. When correlated variables are simultaneously included in the model, XGBoost may distribute their predictive contributions among these variables, resulting in a relatively small individual SHAP value for the steel price index.

Third, the steel price index used in this study is an external market indicator rather than the actual steel procurement price or steel consumption associated with each vessel. The effects of steel demand may already be partially captured by deadweight tonnage and the principal dimensions of the vessel. Moreover, actual steel costs may be influenced by procurement timing, long-term supply contracts, bulk purchasing, inventories, and the time lag between changes in market prices and their transmission to shipbuilding projects. Because ship construction generally extends over a relatively long period, a single steel price index cannot fully represent the actual steel prices paid throughout the entire construction cycle. Consequently, the relatively small SHAP value indicates that the steel price index provides limited additional predictive information after the other variables have been considered, rather than suggesting that steel cost is unimportant in shipbuilding.

In conclusion, when estimating shipbuilding costs, factors such as the energy price index, molded depth, deadweight tonnage, and construction period significantly influence the total cost of ship construction. In contrast, factors like molded breadth and length overall make relatively minor contributions to the output of the estimation model and have a limited impact on shipbuilding costs. Identifying key variables enables shipbuilding enterprises to predict cost trends with greater accuracy, thereby developing more targeted and effective cost-control strategies. Furthermore, understanding these critical

factors allows shipbuilding enterprises to engage in more systematic budget planning and resource allocation, ultimately helping to avoid cost overruns caused by overlooking important variables.

6. Practical Applications

6.1. Examples of Shipbuilding Cost Estimation

To illustrate the best-performing application cases of the PSO-XGBoost model across different vessel types, three test samples with relatively small prediction errors—an 85,000-ton bulk carrier, a refrigerated container ship, and a sawdust carrier—were intentionally selected for presentation. These cases were not randomly selected and are not intended to represent the overall error level of the complete test set. The corresponding cost estimates and absolute errors are presented in Table 8, while the overall predictive performance of the model is evaluated using all test samples in Section 5.3.

Table 8. Actual and estimated costs for selected low-error test cases. Unit: Chinese Yuan (CNY).

Ship Type	Actual Cost Values	Estimated Cost Values	Absolute Value of Error
85,000-ton Bulk carrier	218,837,179.59	218,854,112	16,932.41
Refrigerated container ship	210,654,497.09	210,669,552	15,054.91
Sawdust carrier	204,239,207.96	204,254,432	15,224.04

Source(s): Authors' own work.

As presented in Table 8, the actual cost of the 85,000-ton bulk carrier is 218,837,179.59 yuan, while the estimated cost is 218,854,112 yuan, resulting in an error of 16,932.41 yuan. Similarly, the actual cost of the refrigerated container ship is 210,654,497.09 yuan, with an estimated cost of 210,669,552 yuan, leading to an error of 15,054.91 yuan. For the sawdust carrier, the actual cost is 204,239,207.96 yuan, while the estimated value is 204,254,432 yuan, with an error of 15,224.04 yuan. These three cases represent the test samples with relatively small prediction errors and are presented only to illustrate the best-performing application results of the model. Their errors should not be interpreted as the typical error level of the complete test set or compared directly with the test-set MAPE of 3.87%, which was calculated using all test samples, including those with larger prediction errors.

The selected cases illustrate the procedure through which the trained PSO-XGBoost model generates cost estimates based on the 11 internal and external factors identified in this study. However, these best-performing cases are not used as separate evidence of the model's overall accuracy or generalization capability. The quantitative evaluation of predictive performance is based on the complete independent test set reported in Section 5.3.

6.2. Application and Practical Value

Based on the overall model-performance evaluation presented in Section 5.3 and the application procedure illustrated in Section 6.1, the PSO-XGBoost-based shipbuilding cost estimation model demonstrates potential application value in shipbuilding cost management. By applying this model, key stages in the shipbuilding process—including “pre-planning”, “in-process control”, and “post-assessment”—can be effectively addressed, providing shipbuilding enterprises with accurate and efficient cost estimation support. Specifically, in the pre-planning stage, the model can serve as a critical tool for the rational formulation of ship quotations and the optimization and adjustment of budget preparation. This, in turn, helps shipbuilding enterprises improve decision-making efficiency and enhance market competitiveness. During the in-process stage, the model enables real-time monitoring of shipbuilding cost dynamics, offering robust support for optimizing resource allocation and strengthening cost control. In the post-assessment stage,

the model identifies key drivers and assigns responsibilities for changes in shipbuilding costs, thereby facilitating the establishment of a scientific and transparent cost evaluation system. This ultimately promotes the continuous improvement of cost control mechanisms in shipbuilding enterprises.

(1) Rational development of ship quotations

Before the signing of a contract, shipbuilding enterprises can rely on the PSO-XGBoost-based shipbuilding cost estimation model to obtain real-time cost estimation values and analyze key influencing factors. By inputting internal and external factor data into the model and conducting SHAP analysis, enterprises can gain a precise understanding of the contributions of key factors to shipbuilding costs. This model not only achieves highly accurate cost estimation but also provides scientific, reliable, and interpretable support for contract quotation decisions. After obtaining cost estimation results, shipbuilding enterprises can conduct an in-depth analysis of the potential benefits and risks of undertaking specific projects, taking into account current market conditions and internal development plans. This allows for the formulation of reasonable and competitive quotation strategies in a timely manner. Furthermore, leveraging SHAP analysis and the findings of this study, shipbuilding enterprises are advised to focus on critical parameters such as molded depth and deadweight tonnage. When external variables such as exchange rates or energy price indices experience significant fluctuations, this model can facilitate forward-looking cost estimation, enabling enterprises to promptly develop effective risk mitigation measures. In highly competitive markets, shipbuilding enterprises can adjust their profit margins based on the estimated cost values and the changing trends of key influencing factors to secure orders at more attractive prices. Conversely, when market demand is robust, quotation strategies can incorporate moderate price increases to maximize profit margins. This flexible and dynamic pricing approach not only significantly enhances the efficiency of quotation decision-making but also strengthens the enterprise's ability to respond to market fluctuations. By doing so, shipbuilding enterprises can effectively mitigate risks associated with profit loss or reduced competitiveness caused by cost estimation deviations.

(2) Optimization and adjustment of budget preparation

After signing the contract, shipbuilding enterprises should prepare a comprehensive budget based on the target profit and contract price set by the enterprise. This comprehensive budget encompasses key areas such as revenue, expenses, costs, investment, and cash flow. During the cost budgeting process, the budgeted cost values can be compared with the cost estimates generated by the shipbuilding cost estimation model. By integrating the analysis of key influencing factors, enterprises can further optimize and adjust the relevant budget components to ensure alignment with the strategic objectives.

When the cost budget exceeds the model's estimated value, it may result in resource waste and a decline in competitiveness due to overly conservative planning or improper resource allocation. Drawing on the findings of this study, shipbuilding enterprises should re-evaluate their budgets, with particular emphasis on the construction period as a key influencing factor. If there is potential for optimization in the construction period, it can be shortened by refining construction processes, improving production efficiency, or adopting advanced technologies. These measures can help reduce labor and control costs, thereby enhancing the overall cost structure. Conversely, if the budget falls below the model's estimated value, it may lead to funding shortages and project delays caused by excessive optimism or insufficient consideration of potential risks, such as extended construction periods. In such cases, shipbuilding enterprises should reassess the project schedule and factor in risks, including market fluctuations and extended construction periods.

Additionally, enterprises are advised to increase budget buffers and strengthen their ability to manage construction uncertainties by optimizing construction plans and processes.

When the cost budget aligns closely with the estimated value of the model, it reflects that the budget preparation is scientific, reasonable, and well-suited to the actual situation, with a high level of executability. Shipbuilding enterprises should continue to uphold the rigor of budget preparation and strengthen the monitoring of budget execution. Furthermore, they must ensure that actual costs remain within the budget and systematically summarize lessons learned to provide valuable references for future budget planning. These efforts will collectively enhance the enterprise's overall cost control capabilities.

(3) Real-time dynamic cost control

Shipbuilding enterprises can make timely adjustments and dynamically manage shipbuilding costs based on the key influencing factors identified through SHAP analysis. Drawing on the findings of this study, shipbuilding enterprises can effectively respond to external market fluctuations by implementing proactive strategies. For instance, they can mitigate the impact of price uncertainty on costs by establishing an energy price response mechanism, securing energy and commodity prices through long-term contracts and financial derivatives, and utilizing foreign exchange management tools to hedge against exchange rate risks. Additionally, shipbuilding enterprises can further enhance production efficiency by adopting energy-saving technologies, optimizing project management processes, and improving resource allocation. These measures can help reduce construction costs and shorten construction periods while ensuring that ship performance meets required standards, ultimately achieving the maximization of economic benefits. This systematic cost control strategy not only strengthens the ability of shipbuilding enterprises to withstand market risks but also provides them with sustainable cost advantages in a highly competitive market environment. By integrating advanced cost control measures with risk management tools, shipbuilding enterprises can maintain operational stability, improve profitability, and secure long-term competitiveness. This emphasis on market adaptation and operational stability is also consistent with recent evidence from Chinese transportation equipment manufacturers, showing that technological, organizational, and environmental configurations can jointly contribute to supply chain resilience and sustainable competitiveness (Zhao et al., 2026 [6]).

(4) Scientific and comprehensive cost assessment

After the completion of shipbuilding, cost assessment serves as a core mechanism for shipbuilding enterprises to review the achievement of cost targets and evaluate the effectiveness of cost control efforts. Its systematic implementation is essential for optimizing cost management processes. Drawing on the findings of this study, shipbuilding enterprises can utilize the PSO-XGBoost-based shipbuilding cost estimation model to conduct a comparative analysis between the actual cost incurred during shipbuilding and the cost estimates generated by the model. In addition, SHAP analysis can be employed to quantify the contributions of internal and external factors to shipbuilding costs, thereby uncovering the driving forces behind cost variations. Based on these insights, shipbuilding enterprises can identify the specific factors that result in cost overruns or savings, enabling them to accurately pinpoint the responsible entities or processes. Subsequently, the performance of these entities in managing shipbuilding costs can be systematically evaluated. Through rigorous cost assessment and evaluation, shipbuilding enterprises can address deviations in the cost control process and establish a dynamic feedback mechanism based on assessment results. This enables the timely implementation of positive incentives and corrective measures to address deviations. Moreover, it facilitates the creation of a closed-loop management

cycle of “analysis—attribution—improvement”, driving the reengineering of cost control processes and the optimization of production workflows.

7. Conclusions

This study integrates the Particle Swarm Optimization (PSO) algorithm with the eXtreme Gradient Boosting (XGBoost) algorithm to propose a hybrid PSO-XGBoost model specifically designed for shipbuilding cost estimation. Comparative analysis confirms that the proposed model demonstrates superior estimation accuracy and generalization capability compared to the traditional XGBoost benchmark. Furthermore, validation using real-world datasets indicates that its performance is robust and competitive. These findings underscore the model’s significant advantage in reliably balancing estimation accuracy, making it well-suited to the practical needs of shipbuilding enterprises. Building on the model’s performance, this study identifies 11 key internal and external factors affecting shipbuilding costs. By employing the SHAP interpretable analysis method, the study quantifies the contribution of each factor, significantly enhancing the model’s transparency. The SHAP results reveal that factors such as the energy price index, molded depth, deadweight tonnage, construction period, and USD/CNY exchange rate exert varying degrees of positive or negative influence on final costs. Finally, a case study demonstrates that the PSO-XGBoost model can assist shipbuilding enterprises in achieving accurate and efficient cost estimation. Moreover, it provides a data-driven foundation for optimizing cost control across the entire project lifecycle—encompassing “pre-planning,” “in-process control,” and “post-assessment.”

8. Limitations and Future Research

Although the 11 indirect influencing factors capture the main pathways affecting shipbuilding costs, they may not fully reflect variations associated with labor and major equipment. The dataset used in this study consists of only 51 vessels collected from a single shipyard and mainly covers conventional vessel types constructed over a relatively limited period. This relatively small and concentrated sample may increase the risk of overfitting, limit the generalizability of the proposed model to other shipyards and vessel categories, and increase the variability of the test results. In addition, the use of a single training–test split may make the evaluation results sensitive to the specific samples included in each subset.

Future research may further enrich the set of standardized indirect indicators related to labor and major equipment while retaining the current variable-selection framework. Future studies will expand the vessel database to include larger samples from multiple shipyards, more diverse vessel types, and broader construction periods. Repeated cross-validation and independent external validation will also be employed to further evaluate the robustness and generalization capability of the proposed model.

Author Contributions: Conceptualization, J.Y.; Methodology, G.X.; Validation, Y.S. and G.X.; Formal analysis, Y.S.; Investigation, J.Y.; Data curation, J.Y.; Writing—original draft, J.Y.; Writing—review & editing, Y.S. and G.X.; Visualization, Y.S.; Supervision, G.X.; Project administration, G.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available on request from the corresponding author.

Conflicts of Interest: Author Jing Yang was employed by the Finance Department, Zhongchuan Chengxi Shipyard Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Kong, M.; Ji, R.; Wu, M.; Sun, K.; Zhang, J.; Zhang, Y.; Zhu, R.; Reabroy, R. Correction of sidewall blockage effect for twin-rotor vertical axis tidal stream turbine subjected to different solidities. *Phys. Fluids* **2025**, *37*, 035157. [\[CrossRef\]](#)
- Wu, M.; Ji, R.; Dou, P.; Gao, C.; Zhang, Y.; Zhang, J.; Chen, L.; Fernandez-Rodriguez, E. Energy-harvesting performance of twin-rotor vertical-axis wind turbines with phase interference under different solidities. *J. Mar. Sci. Eng.* **2026**, *14*, 508. [\[CrossRef\]](#)
- Li, X.; Ji, R.; Yang, S.H.; Zhang, Y.; Reabroy, R.; Dou, P.; Chen, L.; Xu, L. A Comprehensive Review of Numerical Simulations on Vortex-Induced Vibration Response Characteristics of Deep-Sea Risers. *Fluids* **2026**, *11*, 159. [\[CrossRef\]](#)
- Zhang, X.; Ji, R.; Sun, K.; Zhang, J.; Zhang, X.; Yin, M.; Kong, M.; Reabroy, R. A review of ocean tidal current energy technology: Advances, trends, and challenges. *Phys. Fluids* **2025**, *37*, 071308. [\[CrossRef\]](#)
- Alblas, G.; Puijn, J. Are current shipbuilding cost estimation methods ready for a sustainable future? A literature review of cost estimation methods and challenges. *Int. Shipbuild. Prog.* **2024**, *71*, 3–28. [\[CrossRef\]](#)
- Zhao, Y.; Su, X.; Wu, F.; Ock, Y.S. Exploring paths for enhancing supply chain resilience in Chinese transportation equipment manufacturers. *Asia Pac. J. Mark. Logist.* **2026**, ahead of print. [\[CrossRef\]](#)
- Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T.Y.; Tegmark, M. KAN: Kolmogorov-Arnold networks[EB/OL]. *arXiv* **2024**, arXiv:2404.19756. [\[CrossRef\]](#)
- Noorizadegan, A.; Wang, S.; Ling, L.; Dominguez-Morales, J.P. A practitioner’s guide to Kolmogorov–Arnold networks. *Comput. Sci. Rev.* **2026**, *62*, 100991. [\[CrossRef\]](#)
- Bodendorf, F.; Franke, J. Synthesis of activity-based costing and deep learning to support cost management: A case study in the automotive industry. *Comput. Ind. Eng.* **2024**, *196*, 110449. [\[CrossRef\]](#)
- Li, L.; Yuan, X.; Wei, S.; Yang, G. Heterogeneous uncertainty matters! Evidence based on firms’ cost management decisions. *Emerg. Mark. Rev.* **2024**, *63*, 101210. [\[CrossRef\]](#)
- Bodendorf, F.; Merkl, P.; Franke, J. Artificial neural networks for intelligent cost estimation—A contribution to strategic cost management in the manufacturing supply chain. *Int. J. Prod. Res.* **2022**, *60*, 6637–6658. [\[CrossRef\]](#)
- Cheng, M.Y.; Cao, M.T.; Mendrofa, A.Y.J. Dynamic feature selection for accurately predicting construction productivity using symbiotic organisms search-optimized least square support vector machine. *J. Build. Eng.* **2021**, *35*, 101973. [\[CrossRef\]](#)
- Kim, S.; Abediniangerabi, B.; Shahandashti, M. Pipeline construction cost forecasting using multivariate time series methods. *J. Pipeline Syst. Eng. Pract.* **2021**, *12*, 04021026. [\[CrossRef\]](#)
- Yang, F.; Tao, F. A bi-objective optimization VRP model for cold chain logistics: Enhancing cost efficiency and customer satisfaction. *IEEE Access* **2023**, *11*, 127043–127056. [\[CrossRef\]](#)
- Liu, T.; Yuan, Y.; Yu, Z. An Intelligent Optimization Control Method for Enterprise Cost under Blockchain Environment. *IEEE Access* **2023**, *11*, 3597–3606. [\[CrossRef\]](#)
- Zhong, W.; Zhong, L.; Dou, Q. The Implementation Strategy of Cost Control and the Construction of a Guarantee Model of Financial BPO in the Cloud Computing Environment. *Int. J. Inf. Syst. Model. Des.* **2025**, *16*, 1–21. [\[CrossRef\]](#)
- Walter, S. AI impacts on supply chain performance: A manufacturing use case study. *Discov. Artif. Intell.* **2023**, *3*, 18. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cremer, L. Underwriting Limits and Optimal Leverage in Commercial Real Estate. *J. Real. Estate Financ. Econ.* **2020**, *60*, 375–395. [\[CrossRef\]](#)
- Li, C.C.; Ni, J. A Continuous-Review Inventory Model: Harnessing the Spot and Futures Price Cointegration for Effective Cost Control. *Omega* **2025**, *137*, 103355. [\[CrossRef\]](#)
- Tambunan, T. A conceptual framework of quality cost chain in strategic cost management. *TQM J.* **2024**, *36*, 762–784. [\[CrossRef\]](#)
- Rounaghi, M.M.; Jarrar, H.; Dana, L.-P. Implementation of strategic cost management in manufacturing companies: Overcoming costs stickiness and increasing corporate sustainability. *Future Bus. J.* **2021**, *7*, 1–8. [\[CrossRef\]](#)
- Pavlatos, O.; Kostakis, H. Exploring the relationship between target costing functionality and product innovation: The role of information systems. *Aust. Account. Rev.* **2022**, *32*, 124–140. [\[CrossRef\]](#)
- Rymarzak, M.; Marmot, A. Higher Education Estate Data Accountability: The Contrasting Experience of UK and Poland. *High. Educ. Policy* **2020**, *33*, 179–194. [\[CrossRef\]](#)
- Zhang, R.; Li, J.L. The application of activity-based costing in the cost calculation of thermal-power enterprise. *Therm. Sci.* **2021**, *25*, 933–939. [\[CrossRef\]](#)
- Ozcalici, M.; Pamucar, D.; Kaya, A.; Tutcu, B.; Karamanci, H.; Gurler, H.E. Strategizing competitive edge in service operations: A model proposal for determining the unit costs through the activity-based cost accounting and the full-consistency method. *Oper. Manag. Res.* **2024**, *17*, 996–1021. [\[CrossRef\]](#)
- Araújo, J.B.C.N.; Souza, A.N.; Joaquim, M.S.; Mattos, L.M.; Junior, I.M.L. Use of the activity-based costing methodology (ABC) in the cost analysis of successional agroforestry systems. *Agrofor. Syst.* **2020**, *94*, 71–80. [\[CrossRef\]](#)
- Duarte, A.H.; Barbalho, S.C.M.; Vieira, D.; Bravo, A. A quantitative and qualitative study of life cycle costing in defense projects and programs. *Manag. Rev. Q.* **2024**, *75*, 2207–2233. [\[CrossRef\]](#)

28. Lu, J.; Wu, D.; Dong, J.; Dolgui, A. A decision support method for credit risk based on the dynamic Bayesian network. *Ind. Manag. Data Syst.* **2023**, *123*, 3053–3079. [\[CrossRef\]](#)
29. He, M.; Chen, A.; Han, Y.; He, C.; Geng, X. Effects of chief economists in financial institutions on capital market development. *Int. Rev. Financ. Anal.* **2025**, *105*, 104403. [\[CrossRef\]](#)
30. Jiang, L.; Wang, H.; Xiao, Y.; Xu, L.; Chen, H. Exploring the association between volatile organic compound exposure and chronic kidney disease: Evidence from explainable machine learning methods. *Ren. Fail.* **2025**, *47*, 2520906. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Ting, T.; Aslam, M.M.; Imran, M.H.; Wah, K.K. Predicting the financial performance of microfinance institutions with machine learning techniques. *J. Model. Manag.* **2025**, *20*, 322–347. [\[CrossRef\]](#)
32. Kim, S.; Makhmalbaf, A.; Shahandashti, M. Forecasting the architecture billings index (ABI) using machine learning predictive models. *Eng. Constr. Archit. Manag.* **2024**, *32*, 5371–5393. [\[CrossRef\]](#)
33. Yan, H.; He, Z.; Gao, C.; Xie, M.; Sheng, H.; Chen, H. Investment estimation of prefabricated concrete buildings based on XGBoost machine learning Algorithm. *Adv. Eng. Inform.* **2022**, *54*, 101789. [\[CrossRef\]](#)
34. Huang, J.; Huang, W.; Quan, W.; Xing, Y. Hybrid Intelligent Model for Estimating the Cost of Huizhou Replica Traditional Vernacular Dwellings. *Buildings* **2024**, *14*, 2623. [\[CrossRef\]](#)
35. Paik, Y.; Chung, F.; Ashuri, B. Preliminary Cost Estimation of Pavement Maintenance Projects through Machine Learning: Emphasis on Trees Algorithms. *J. Manag. Eng.* **2025**, *41*, 04025027. [\[CrossRef\]](#)
36. Zhang, J.; Yuan, J.; Mahmoudi, A.; Ji, W.; Fang, Q. A data-driven framework for conceptual cost estimation of infrastructure projects using XGBoost and Bayesian optimization. *J. Asian Archit. Build. Eng.* **2025**, *24*, 751–774. [\[CrossRef\]](#)
37. Deepa, G.; Niranjana, A.J.; Balu, A.S. A hybrid machine learning approach for early cost estimation of pile foundations. *J. Eng. Des. Technol.* **2025**, *23*, 306–322. [\[CrossRef\]](#)
38. Hammann, D. Big data and machine learning in cost estimation: An automotive case study. *Int. J. Prod. Econ.* **2024**, *269*, 109137. [\[CrossRef\]](#)
39. Srinadhraju, S.; Mishra, S.; Satapathy, C.S. Software Cost Estimation Using Social Group Optimization. *Comput. Syst. Sci. Eng.* **2024**, *48*, 1641–1668. [\[CrossRef\]](#)
40. Draz, M.M.; Emam, O.; Azzam, S.M. Software cost estimation predication using a convolutional neural network and particle swarm optimization algorithm. *Sci. Rep.* **2024**, *14*, 13129. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Fafandjel, N.; Zamarin, A.; Hadjina, M. Shipyard production cost structure optimisation model related to product type. *Int. J. Prod. Res.* **2010**, *48*, 1479–1491. [\[CrossRef\]](#)
42. Jiang, L.; Strandenes, S.P. Assessing the cost competitiveness of China's shipbuilding industry. *Marit. Econ. Logist.* **2012**, *14*, 480–497. [\[CrossRef\]](#)
43. Broderick, D.; Wright, P.; Kattan, R. Minimizing the Cost of Coating Ships. *J. Ship Prod. Des.* **2013**, *29*, 142–149. [\[CrossRef\]](#)
44. Carberry, S. The Guessing Game of Shipbuilding Costs. *Natl. Def.* **2024**, *108*, 14.
45. Tran, K.N.; Haralambides, H.; Notteboom, T.; Cullinane, K. The costs of maritime supply chain disruptions: The case of the Suez Canal blockage by the 'Ever Given' megaship. *Int. J. Prod. Econ.* **2025**, *279*, 109464. [\[CrossRef\]](#)
46. Alblas, G.; Pruyn, J. A Novel Concept for Steel Building Cost Estimation in Shipbuilding Using Process and Product Similarities. *J. Mar. Sci. Eng.* **2024**, *12*, 1936. [\[CrossRef\]](#)
47. Gualeni, P.; Vairo, T. A prediction tool for maintenance costs estimation during the design process of a ship engine room. *J. Ocean Eng. Mar. Energy* **2023**, *9*, 653–663. [\[CrossRef\]](#)
48. Li, J.; Sun, M.; Han, D.; Wang, J.; Mao, X.; Wu, X. A knowledge discovery and reuse method for time estimation in ship block manufacturing planning using DEA. *Adv. Eng. Inform.* **2019**, *39*, 25–40. [\[CrossRef\]](#)
49. Lin, C.-K.; Shaw, H.-J. Feature-based estimation of preliminary costs in Shipbuilding. *Ocean Eng.* **2017**, *144*, 305–319. [\[CrossRef\]](#)
50. Azhar, A.; Kristiyono, T.A. Method for Estimating Cost of New Ship Building with Linear Regression. *IOP Conf. Ser. Earth Environ. Sci.* **2022**, *972*, 012014. [\[CrossRef\]](#)
51. Meng, Y.; Xu, C.; Li, T.; Liu, T.; Tang, L.; Zhang, J. Prediction of Total Organic Carbon Content in Shale Based on PCA-PSO-XGBoost. *Appl. Sci.* **2025**, *15*, 3447. [\[CrossRef\]](#)
52. Li, D.; Zhu, D.; Tao, T.; Qu, J. Power Generation Prediction for Photovoltaic System of Hose-Drawn Traveler Based on Machine Learning Models. *Processes* **2023**, *12*, 39. [\[CrossRef\]](#)
53. Xing, W.; Gao, H.; Qiao, X. Measurement of gas volume fraction in gas-liquid two-phase flow using arrayed fiber-optic probes combined with the PSO-BP-AdaBoost algorithm. *Opt. Fiber Technol.* **2025**, *93*, 104264. [\[CrossRef\]](#)
54. An, J.; Guo, W.; Lv, T.; Zhao, Z.; He, C.; Zhao, H. Joint Prediction of the State of Charge and the State of Health of Lithium-Ion Batteries Based on the PSO-XGBoost Algorithm. *Energies* **2023**, *16*, 4243. [\[CrossRef\]](#)
55. Abedin, J.; Franklin, F.; Mahmud, I.M.S. A Two-Stage Optimisation of Ship Hull Structure Combining Fractional Factorial Design Technique and NSGA-II Algorithm. *J. Mar. Sci. Eng.* **2024**, *12*, 411. [\[CrossRef\]](#)
56. Su, M.; Su, Z.; Bae, H.S.; Li, J.; Park, K.-S. An exploration of shipbuilding price prediction for container ships: An integrated model application of deep learning. *Res. Transp. Bus. Manag.* **2025**, *59*, 101248. [\[CrossRef\]](#)

57. Xue, Y.; Zhang, X. Digital transformation and corporate capital structure: Evidence from China. *Pac. Basin Financ. J.* **2024**, *84*, 102299. [[CrossRef](#)]
58. Liu, B.; Lin, H.; Chen, Y.; Yang, C. Prediction of Rock Unloading Strength Based on PSO-XGBoost Hybrid Models. *Materials* **2024**, *17*, 4214. [[CrossRef](#)] [[PubMed](#)]
59. Yu, Z.; Li, C.; Zhou, J. Tunnel Boring Machine Performance Prediction Using Supervised Learning Method and Swarm Intelligence Algorithm. *Mathematics* **2023**, *11*, 4237. [[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.