Scientific Journal of Informatics
Vol. 12, No. 2, May 2025

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Freshwater Filling Optimization Based on Price Using XGBoost and Particle Swarm Optimization on Cargo Ship Voyage

Ilham Yulianto[1], Muhammad Dzulfikar Fauzi[2*], Pima Hani Safitri[3]
[1, 2, 3]Department of Informatics, Telkom University Surabaya, Indonesia

Abstract.
Purpose: Efficient freshwater management is critical in cargo ship operations, yet current practices often involve fixed refilling strategies that ignore price differences across ports and fail to predict actual consumption accurately. These inefficiencies lead to unnecessary operational costs. To address this, the study introduces a combined approach using XGBoost for predict freshwater usage and Particle Swarm Optimization (PSO) to minimize refilling costs through optimal port selection.
Methods: Freshwater demand was predicted using an XGBoost regression model trained on real operational data from 2024, which included historical voyage distances and freshwater consumption records from cargo ships. Based on these predictions, Particle Swarm Optimization (PSO) was applied to identify cost-efficient refilling locations along each ship's route, minimizing total water procurement cost while satisfying operational constraints. The proposed framework was validated through simulated voyage scenarios to evaluate its impact on cost efficiency and planning effectiveness.
Result: The integration of XGBoost and PSO effectively optimized freshwater refilling strategies, achieving a relative prediction error of 9.48% in freshwater consumption prediction and cost savings from 9 to 40% from across 3 ships sample through strategic port selection based on consumption patterns and price variability.
Novelty: Unlike prior works focused on fuel or generic logistics optimization, aim of this study is to combine XGBoost and PSO for optimizing freshwater refilling on cargo ship voyages using actual operational data. The results demonstrate practical, scalable improvements in cost efficiency, making a novel contribution to maritime resource planning.

Keywords: Freshwater management, Operational efficiency, XGBoost, Particle swarm optimization
Received May 2025 / **Revised** June 2025 / **Accepted** June 2025

INTRODUCTION

In the pursuit of greater operational efficiency, the maritime industry is increasingly prioritizing the strategic management of onboard resources, freshwater being one of the most essential yet often overlooked [1], [2], [3], [4]. Freshwater plays a vital role in daily ship operations, supporting crew welfare, engine cooling, equipment maintenance, and overall voyage sustainability [5], [6]. However, the conventional approach to freshwater refilling tanks at the start of a voyage without considering price variations at different ports leads to unnecessary costs and operational inefficiencies. These issues are compounded by limited tank capacity, which is further constrained by ship draft regulations and the need to balance cargo and fuel loads.

Amid growing cost of resources and tightening environmental regulations now demand more intelligent and data-driven resource planning [7], [8]. As part of broader energy efficiency strategies such as the Ship Energy Efficiency Management Plan (SEEMP), freshwater optimization is becoming critical for reducing operational costs while enhancing sustainability [9], [10], [11]. Despite its importance, freshwater management is rarely treated with the same level of strategic planning as fuel or cargo logistics, making it a valuable target for innovation through predictive and optimization technologies.

Currently, Freshwater management on cargo ships is often inefficient. They rely on conventional practices like filling tanks at the voyage's start without considering water price variations at different ports. This overlooks cost-saving opportunities when cheaper water is available en route. Moreover, freshwater tank capacity is limited by ship draft constraints, and more cargo or fuel reduces available water storage, making

*Corresponding author.
Email addresses: ilhamyulianto@student.telkomuniversity.ac.id (Yulianto),
dzulfikarf@telkomuniversity.ac.id (Fauzi)*, phanisafitri@telkomuniversity.ac.id (Safitri)
DOI: [10.15294/sji.v12i2.24988](https://doi.org/10.15294/sji.v12i2.24988)

it crucial to optimize freshwater use throughout the journey [12]. In cargo ship operations, three main resources are loaded: cargo, fuel, and freshwater. Cargo takes the highest priority as it drives profitability, followed by fuel, which is essential for powering the ship. Freshwater, while ranked third, is vital for all water supply needs on cargo ships such as crew consumption and engine cooling [13], [14]. Efficient freshwater management is critical, especially on long voyages, helping minimize waste and free up space for other essential loads. A voyage refers to the ship's journey from the departure port to the destination, encompassing route planning, cargo handling, documentation, travel time, transit points, operational costs, and external factors like weather and sea conditions. [15], [16].

The research underscores the importance of careful tank allocation to maintain operational efficiency, particularly on general cargo ships where balancing fuel and freshwater capacities is crucial to stay within safe load limits. Optimizing freshwater usage not only frees up valuable capacity for cargo and fuel but also reduces costs and ensures sufficient supply for essential operations throughout the voyage [14], [17]. Applying machine learning, particularly XGBoost regression, offers an effective solution for optimizing freshwater management on general cargo ships by integrating operational data with external factors [18], [19], [20]. Efficient tank allocation is crucial for balancing fuel, freshwater, and cargo loads within safe limits. Optimizing freshwater usage frees up capacity, cuts costs, and ensures reliable supply during voyages. Studies show XGBoost outperforms traditional models, achieving low MAPE values (15.6%–22.9% for wave height, 8.3%–13.4% for wave period), reduced biases (−2.56% to −10.61%), and remarkable computational speed, completing two years of predictions in just 0.03 seconds on a single CPU, demonstrating its strong potential for shipboard freshwater optimization [21]. And for added example, in a multivariate ensemble model predicting community water consumption, XGBoost demonstrated strong performance in handling complex time series data. Combined with SARIMAX and Prophet, the ensemble forecasted up to 36 months ahead with results of 360.075 MSE, 18.976 RMSE, 15.814 MAE, and 0.027 MAPE, highlighting XGBoost's effectiveness within advanced forecast and predicting frameworks [22].

Optimizing the selection of freshwater refill locations based specifically on the lowest prices can be done using the Particle Swarm Optimization (PSO) method. Studies have shown that the Particle Swarm Optimization (PSO) algorithm is effective for determining optimal locations in water resource management due to its ability to handle optimization problems with many variables and non-linear relationships, in various cases, PSO can successfully applied under a 50% reliability constraint for multiple future scenario, and PSO proven can help to increase a decision tree model performance in location selection model from 96,53% without PSO and 97,78% with PSO [23], [24], [25].

While operational optimization of cargo ships has been widely explored, freshwater management, particularly using machine learning, remains under-researched, especially in Indonesia. This study aims to develop a freshwater management optimization model using XGBoost regression and Particle Swarm Optimization (PSO), based on real-world voyage data. The goal is to improve freshwater usage efficiency, reduce operational costs, and support better decision-making for onboard freshwater refilling.

METHODS

Dataset and research flow

The initial stage of this study involved collecting operational ship data required for model development. The dataset comprises structured records from PT Salam Pacific Indonesia Lines (SPIL), covering real-world cargo ship operations throughout 2024. The data was gathered during an internship program conducted by the first author at SPIL from August to December 2024.

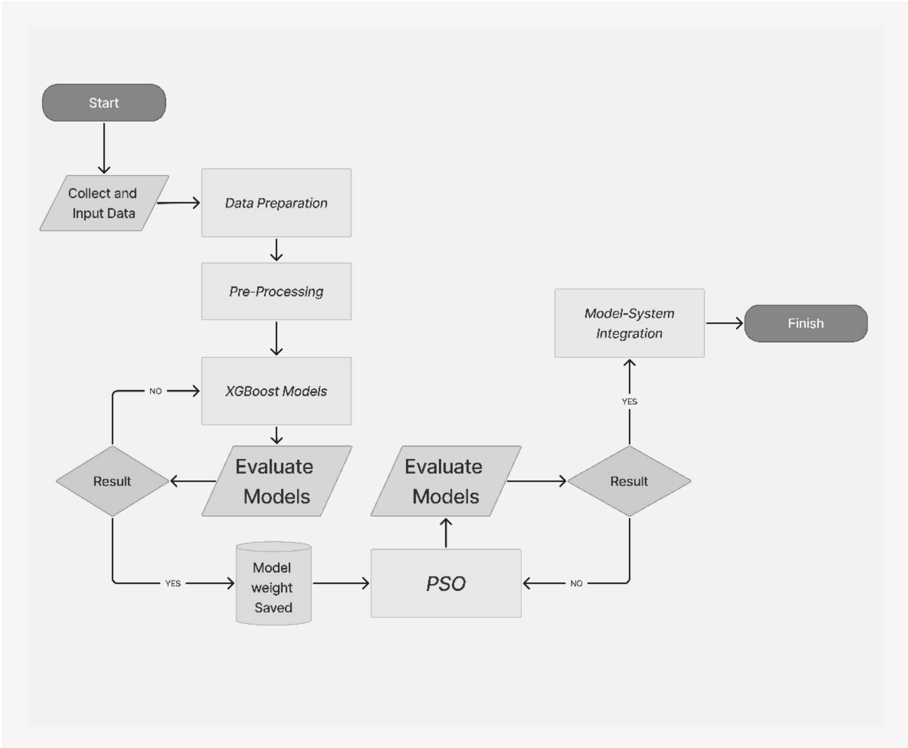

Figure 1. Research flowchart

Figure 1 illustrates the overall research workflow undertaken in this study. The process begins with the collection of operational ship data, the dataset includes:

1. Voyage details: port sequences and trip segments,
2. Ship specifications: tank capacity, gross tonnage, and deadweight tonnage,
3. Distance data: nautical miles between each port pair,
4. Freshwater consumption: historical usage in liters per route,
5. Freshwater price: port-specific cost per liter across 38 Indonesian ports.

Following this, a data preprocessing phase is conducted, where categorical variables such as ship routes are transformed using one-hot encoding, and numerical features are standardized where appropriate. The cleaned data is then split into training and testing sets for the development of the XGBoost regression model, which predicts freshwater consumption based on voyage distance. The output from this model serves as input for the Particle Swarm Optimization (PSO) algorithm, which determines the optimal freshwater refill points and quantities by minimizing total cost while satisfying operational constraints, the tank capacity. Model evaluation and analysis are using MSE and RMSE for both models (XGBoost and PSO) performed to assess prediction accuracy and quantify cost savings achieved through the optimized refilling strategy.

For data collection, the first data collected is raw data derived from ship operational records, where the raw data obtained generally only contains operational data obtained in a certain period, and not directly in large quantities.

Table 1. Destination and freshwater usage raw data

	Destination	Date at sea		At Sea	Arrival	Fresh water cons. At sea
OKI	JAKARTA	01/01/2024 14:21:00	01/02/2023 14:18:00	92/45,5979166667	58/44,5958333333	8000
JAKARTA	SURABAYA	01/03/2024 13:31:00	01/04/2023 12:19:00	52/45,5631944444	17/45,5131944444	8000
SURABAYA	BALIKPAPAN	01/05/2024 13:54:00	01/07/2023 11:44:00	13/45,5791666667	08/45,4888888889	13000
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
SURABAYA	MAKASSAR	02/02/2024 14:16:00	02/04/2023 13:23:00	24/45,5944444444	18/45,5576388889	15000
MAKASSAR	BAU-BAU	02/04/2024 14:10:00	02/05/2023 13:45:00	84/45,5902777778	48/45,5729166667	8000
BAUBAU	KENDARI	02/07/2024 13:00:00	02/08/2023 12:04:00	75/45,5416666667	40/45,5027777778	7000

The raw data refers to the initial unprocessed information that will be aggregated from all similar datasets. Preliminary preprocessing will then be conducted to remove missing values and anomalies, such as input results that deviate significantly from the majority of the data.

Using the example of the raw data above, in Table 1, such raw data will be consolidated with other similar datasets, resulting in a final dataset that retains only the 'departure and destination (trips)' and 'freshwater consumption' columns.

Table 2. Route and distance raw data

ROUTE	CODE	DISTANCE
OKI-JAKARTA	1	330
JAKARTA-SURABAYA	2	438
SURABAYA-BALIKPAPAN	3	600
.	.	.
.	.	.
.	.	.
MAKASSAR-BAUBAU	7	244
BAUBAU-KENDARI	8	248
KENDARI-SURABAYA	9	1143,68

Additionally, there is another type of raw data, as illustrated in Table 2, which contains information on trips from Port A to Port B, along with the corresponding distance measured in nautical miles. This type of raw data will be utilized to derive the distance data, which will be further elaborated in the following section.

In addition to the examples of raw data previously presented in Tables 1 and 2, other collected datasets may not follow the same raw data format, as some originate directly from the operational records of vessels owned by PT. Salam Pacific Indonesia Lines (SPIL). Once all relevant raw and operational data required to support this study have been gathered, the resulting processed dataset ready for further analysis, usage and modelling, can be observed, as shown in Table 3 below. Table 3 shows the Ship route, distance, and freshwater usage data. This data amounts to 1186 samples, with the farthest Distance at 1838 Nautical miles, and the highest Freshwater usage is 19,200 liters.

Table 3. Ship route, distance, and freshwater usage

Ship Route	Freshwater	Distance (NM)
REDE-SURABAYA	100	17
SURABAYA-REDE	100	17
DOCKING-SURABAYA	200	50
.	.	.
.	.	.
.	.	.
SURABAYA-MANOKWARI	17900	1397
MANOKWARI-SURABAYA	18000	1648
JAYAPURA-SURABAYA	19200	1838

The preprocessed dataset serves as the primary input for training the XGBoost regression model, which is designed to predict freshwater consumption based on voyage distance and route information. Prior to training, categorical variables, particularly ship routes, are transformed using one-hot encoding to ensure compatibility with the model, while numeric features such as distance are retained in their original format.

One-hot encoding is necessary for categorical data like Ship route because machine learning models such as XGBoost cannot interpret textual labels directly. Assigning arbitrary numeric values (e.g., 1 for "REDE–SURABAYA", 2 for "SURABAYA–REDE") would imply a false ordinal relationship between categories, which doesn't exist and could mislead the model. One-hot encoding solves this by creating a separate binary column for each unique route, allowing the model to treat each route independently without implying any order or priority. In contrast, Distance is already a numeric, continuous feature that naturally carries quantitative meaning, so it can be used directly without encoding.

This combination of features, numerical distance values and one-hot encoded categorical routes, ensures the dataset is properly structured for optimal model performance. The processed dataset is then used to train the XGBoost regression model to predict freshwater consumption based on voyage characteristics.

To support the development of the predictive and optimization models, additional structured data is provided in Tables 4 and 5. Table *4 shown* detailed specifications of the vessels (ASJ, REN, AKA) used in the case study, such as tank capacity, gross tonnage, and deadweight tonnage.

Table 4. Ship operational data

ship	tank_capacity	gross tonnage	deadweight tonnage
ASJ	28000	6093	8528
REN	25000	5823	8318
AKA	27000	5501	8037

The three used vessel in this study (ASJ, REN, AKA) each have freshwater tank capacities ranging from 25,000 to 28,000 liters. This capacity classifies them within the feedermax to small panamax categories, typically used for regional or short intercontinental routes [14][26], [27], [28], [29]. Based on this classification, their operational profiles generally include shorter voyages (regional inter-island voyages), smaller crew complements (typically 10–15 personnel), and more frequent freshwater resupply opportunities compared to larger ocean-going vessels.

Given these characteristics, the selected ships represent the lower end of the medium-sized cargo fleet, where freshwater storage is inherently constrained. These constraints arise not only from tank size limitations but also from the need to balance other critical loads, such as cargo volume, fuel capacity, and ship draft requirements, as outlined in the Introduction. Therefore, optimizing freshwater refilling under these practical constraints provides a realistic and operationally relevant test case for the proposed model. Table 5 shown the freshwater prices across 38 Indonesian ports as of 2024, serving as key input variables in the PSO algorithm to identify cost-effective refilling strategies.

Table 5. Freshwater price

LOCATION	COST
AMBON	40.000
BALIKPAPAN	24.000
BANJARMASIN	35.000
.	.
.	.
.	.
PADANG	37.000
PALARAN	17.500
PEKAN BARU	75.000

XGBoost model

The next step is to split the train and test data with a ratio of 8:2, then proceed to the XGBoost regression model with 7 different configurations, shown in Table 6.

Table 6. XGBoost configuration table

	Learning rate	Seeds	Boost rounds	Early stopping round	Max depth
Config 1	0,1	100	10.000	50	8
Config 2	0,1	200	12.000	50	8
Config 3	0,15	100	12.000	50	8
Config 4	0,15	200	15.000	50	8
Config 5	0,2	100	10.000	50	8
Config 6	0,2	200	15.000	50	8
Config 7	0,25	200	15.000	50	8

To develop an effective XGBoost regression model for predicting freshwater consumption, seven configurations were systematically evaluated by varying key hyperparameters: learning rate, number of boosting rounds, and random seed, while holding other parameters constant.

The hyperparameters presented in Figure 4 were selected to explore the impact of learning rate, number of boosting rounds, and random seed on model performance, specifically in the context of freshwater consumption prediction. The learning rate controls how quickly the model adapts to the training data; lower values are more conservative and may prevent overfitting but require more rounds, while higher values enable faster convergence at the risk of overshooting. The number of boosting rounds determines how many additive decision trees are built; more rounds allow the model to capture more complex patterns but increase computation time. The random seed influences the initial state and data shuffling, which can affect reproducibility and variability across runs. These configurations are consistent with widely recognized best practices in XGBoost modeling, as demonstrated in multiple prior studies [30], [31].

Learning rate was incrementally increased from 0.1 to 0.25 to assess the model's sensitivity to convergence speed and performance. This parameters use to controls the step size at each iteration to balance between training speed and model accuracy; a lower value typically prevents overfitting but requires more boosting rounds [32], [33]. Initial configurations with lower rates served as baselines, while higher rates were paired with increased boosting rounds (10,000 to 15,000) which refers to a single iteration of training and new decision tree and adding it to existing ensemble of trees to incrementally improving model performance by sequentially reducing the redisual error.

Then MSE (Mean Squared Error) and RMSE (Root Mean Squared Error) used to evaluate the model. MSE and RMSE are widely used and well-suited evaluation metrics for XGBoost in regression tasks because they align closely with the model’s default objective and optimization process [34], [35], [36], [37]. XGBoost, when used for regression, typically minimizes squared loss during training, which is exactly what MSE represents the average of the squared differences between predicted and actual values. RMSE, being the square root of MSE, brings the error metric back to the same unit as the target variable (e.g., liters of water), making it easier to interpret. These metrics penalize larger errors more than smaller ones, which is desirable in many practical applications where significant deviations are costlier [38], [39], [40].

Particle swarm optimization (PSO)
Then, PSO model is the standard PSO with inertia weight. It updates particle velocities based on inertia, cognitive (personal best), and social (global best) components, with parameters set as: 50 particles, 100 iterations, inertia weight $\boldsymbol{w} = \mathbf{0.5}$, and learning factors $\mathbf{c1 = c2 = 1{,}5}$. In this case, PSO is applied to find the cheapest and best position for a ship to replenish freshwater during a voyage, where the port distances in the route have been predicted by an XGBoost model. PSO also determines the optimal volume of freshwater to refill at each port, ensuring cost efficiency while satisfying operational constraints.

XGBoost and PSO model integration for optimization
After the XGBoost and PSO models are finalized, the integration process begins by utilizing voyage data, such as the example provided in Table 7, which outlines port sequences and corresponding trip segments for each vessel.

Table 7. Ship voyages example

VOYAGE	PORTS	TRIPS
2-23	SURABAYA;MAKASSAR;BAUBAU;KENDARI	SURABAYA-MAKASSAR;MAKASSAR-BAUBAU;BAUBAU-KENDARI;KENDARI-SURABAYA

Shown in Table 7 is one example of the voyage of the ASJ vessel, it has ‘PORTS’ which means which ports will be visited during the voyages. Then 'TRIPS' describes how the ship's route (voyage) from the starting port, going to several destination ports to deliver and transport containers, to return to the starting port again. Then to process how much freswater is needed in 'TRIPS', each trip in the voyage will be broken down, separated based on its semicolon, like this ['SURABAYA-MAKASSAR', 'MAKASSAR-BAUBAU', 'BAUBAU-KENDARI', 'KENDARI-SURABAYA']. Then for each trip will be adjusted to the list of DISTANCE in the Table 1. Result: [[416], [1368], [248], [757]]. And using the XGBoost model, it will predict how much freshwater (in liters) will be used for each trip in the voyage.

Prior to the PSO optimization process, freshwater prices at each port along the voyage route are identified and matched to their respective locations. This step ensures that the optimization algorithm has access to accurate cost data for evaluating refilling options across different segments of the journey. As an example using the same voyages data, the PORTS are obtained as follows, ['SURABAYA', 'MAKASSAR', 'BAUBAU', 'KENDARI'], And water price data for each city will be obtained based on the water prices in Table 2, as follows, [35000, 30000, 31000, 40000].

After all distances have predicted their freshwater needs, and the list of freshwater prices at each port in the voyages is also available, then Particle Swarm Optimization (PSO) algorithm is designed to identify the most cost-effective freshwater refill strategy by considering two primary factors: (1) minimizing total freshwater procurement cost, and (2) ensuring that the ship’s freshwater supply remains sufficient to complete the entire voyage, including a safe return to the starting port based on the desired voyage.

To achieve this, PSO searches for refill locations by first prioritizing ports with the lowest unit prices of freshwater encountered along the route. For each candidate solution (i.e., refill scenario), PSO evaluates whether the cumulative freshwater supplied at selected ports will meet or exceed the predicted consumption for all voyage segments, as predicted by the XGBoost model. While no explicit weight coefficients are used, the objective function implicitly penalizes infeasible solutions (e.g., those resulting in water shortage before journey completion), guiding particles toward options that are both cost-efficient and operationally valid. Additionally, PSO explores alternative refill combinations along the route, enabling the model to discover near-optimal solutions that balance early low-cost refills with operational constraints like tank capacity and port spacing. This ensures that the vessel arrives at its final destination with minimum cost and without violating supply requirements.

RESULTS AND DISCUSSIONS

The dataset used for this study consists of 1,186 records, each containing a categorical ship route, travel distance (in nautical miles), and the target variable freshwater consumption (in liters). The consumption values span a broad range, from 100 liters to 19,200 liters, illustrating significant variance in provisioning needs.

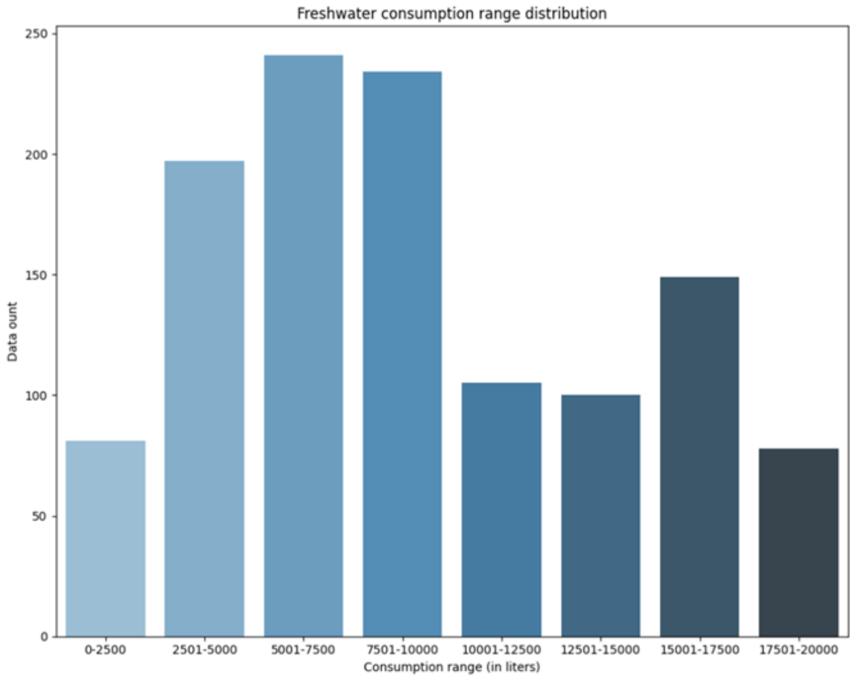

Figure 5. Freshwater consumption dataset distribution

As shown in Figure 2, the distribution is right-skewed, with most values concentrated between 5,000 and 10,000 liters, and a substantial number of observations also found in both lower (0–2,500 liters) and higher (15,000–17,500 liters) ranges.

XGBoost result

This variability and skewness motivated the use of XGBoost regression, a model well-suited for handling non-linear patterns and mixed type features. To optimize performance, seven configurations were tested with varying learning rates (0.1 to 0.25), boosting rounds (10,000 to 15,000), and seed values, while maintaining constant maximum depth and early stopping rounds. These hyperparameter choices were designed to balance learning stability with the model's capacity to capture complex patterns in the data.

Table 8 displays RMSE and MSE result from each of 7 configurations.

Table 8. RMSE and MSE results

	RMSE	MSE
Config 1	762.76	581814.58
Config 2	762.76	581814.58
Config 3	764.05	583777.25
Config 4	764.05	583777.25
Config 5	761.39	579729.03
Config 6	761.39	579729.03
Config 7	758.44	575235.31

The performance metrics, RMSE and MSE for each configuration are summarized in Table 8. The lowest RMSE was 758.44 liters (Config 7), with an associated MSE of 575,235.31. The reason this configuration have performed best is the balance it strikes between learning speed and model complexity. A learning rate of 0.25 allows the model to converge faster while still avoiding overfitting, especially when paired with a sufficiently high number of boosting rounds (15,000), which enables deeper learning over many iterations. In contrast, the highest RMSE recorded was 764.05 liters (Configs 3 and 4). The RMSE provides a direct interpretation of average prediction error in the same unit as the target variable (liters).

To interpret the error in percentage terms, compare RMSE to the range of actual values. Using the average freshwater consumption (approx. 8,000 liters, based on the distribution histogram), the RMSE of 758.44 liters corresponds to a relative error of:

$$\text{Relative Error (\%)} = \frac{758.44}{8{,}000} \times 100 \approx 9.48\%$$

This indicates that the model's predictions, on average, deviate from the actual values by less than 10%, which is considered reasonable given the wide data spread and high variability across routes. Then visualization on how differece between actual and predicted values in Figure 3.

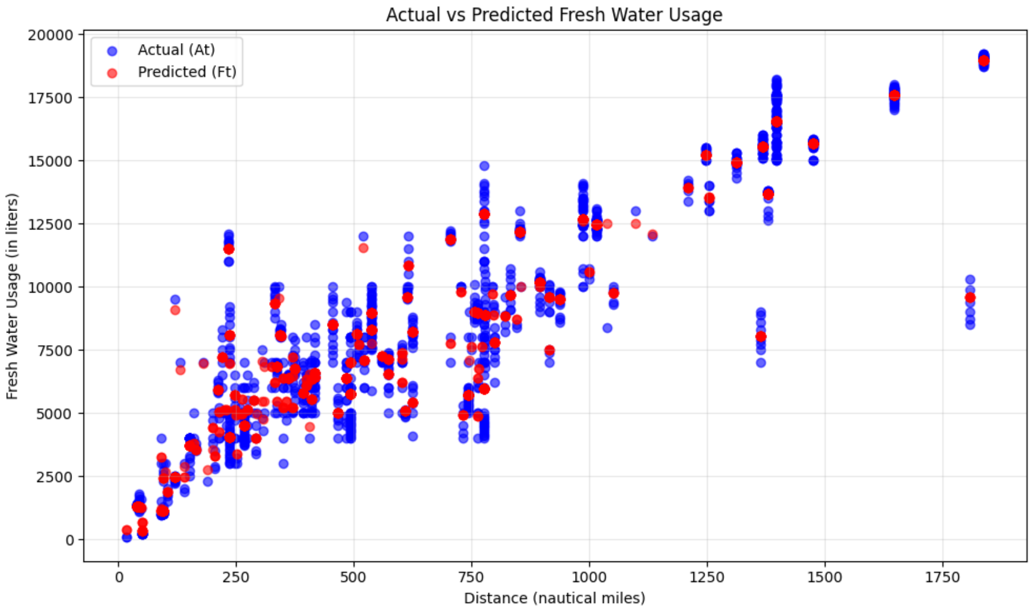

Figure 2. Actual and predicted freshwater consumption chart

The scatter plot comparing actual and predicted fresh water usage visually confirms the model's predictive accuracy. The predicted values closely align with actual observations across all distances, with only minor deviations. This visual consistency reinforces the previously calculated relative error of approximately 9.48%, indicating that the XGBoost model reliably captures the relationship between distance and water consumption.

Beside of it, the model's relative prediction error of 9.48% also indicates that, on average, the predicted freshwater consumption deviates by less than 10% from actual usage across various voyages. In the context of ship operations, this level of accuracy is operationally acceptable, particularly for medium-sized vessels like the feedermax to small panamax classes analyzed in this study. These vessels typically carry freshwater tank capacities between 25,000 to 35,000 liters, meaning a 9.48% error translates to an absolute deviation of roughly 2,400 to 2,700 liters.

PSO result

For the PSO results, voyage data from the three vessels listed will be used. Three sample voyages will be selected from each vessel (ASJ, REN, and AKA) to serve as PSO result samples. Table 9 presents 3 voyages from ASJ ship.

Table 9. ASJ voyage

	ASJ Ship Voyage
1	SURABAYA-MAKASSAR;MAKASSAR-BAUBAU;BAUBAU-KENDARI;KENDARI-SURABAYA
2	SURABAYA-MAKASSAR;MAKASSAR-KENDARI;KENDARI-BAUBAU;BAUBAU-SURABAYA
3	SURABAYA-MAKASSAR;MAKASSAR-MANOKWARI;MANOKWARI-NABIRE;NABIRE-BIAK;BIAK-SERUI;SERUI-AMBON;AMBON-SURABAYA

Then, Table 10 shows the cost results for the freshwater filling optimization for each voyage, along with a comparison to the scenario without using PSO.

Table 10. ASJ PSO optimization result

Voyage 1			
	TRIPS	No-PSO (Freshwater)	PSO (Freshwater)
N1	SURABAYA-MAKASSAR	28.000	6.412
N2	MAKASSAR-BAUBAU	6.722	28.000
N3	BAUBAU-KENDARI	0	310
N4	KENDARI-SURABAYA		
Total Cost in Rupiah (in thousand scale)		1.181.660	1.074.030
Difference	9.1%		
Voyage 2			
N1	SURABAYA-MAKASSAR	28.000	6.412
N2	MAKASSAR-KENDARI	0	28.000
N3	KENDARI-BAUBAU	0	0
N4	BAUBAU-SURABAYA	6.994	492
Total Cost in Rupiah (in thousand scale)		1.196.660	1.079.672
Difference	9.74%		
Voyage 3			
N1	SURABAYA-MAKASSAR	28.000	6.605
N2	MAKASSAR-MANOKWARI	0	19.670
N3	MANOKWARI-NABIRE	0	875
N4	NABIRE-BIAK	0	487
N5	BIAK-SERUI	0	0
N6	SERUI-AMBON	21.102	21.362
N7	AMBON-SURABAYA	0	100
Total Cost in rupiah (in thousand scale)		1.763.540	1.558.205
Difference	13.4%		

In ASJ ship case, PSO optimizes freshwater refills by minimizing initial high-cost purchases at Surabaya (35,000/liters) and strategically allocating refills in subsequent ports. This approach reduces costs by 9.1% to 13.4% across different routes by maintaining necessary freshwater levels while avoiding excessively high-cost refills, especially when lower-cost options become available later in the route.

Next, Table 11 presents 3 voyages from REN ship.

Table 11. REN voyages

	REN Voyages
1	JAKARTA-SAMARINDA;SAMARINDA-BALIKPAPAN;BALIKPAPAN-JAKARTA
2	JAKARTA-SAMARINDA;SAMARINDA-BALIKPAPAN;BALIKPAPAN-SEMARANG;SEMARANG-JAKARTA
3	JAKARTA-BANJARMASIN;BANJARMASIN-SAMARINDA;SAMARINDA-SURABAYA;SURABAYA-JAKARTA

Table 12 shows the cost results for the freshwater filling optimization for each voyage, along with a comparison to the scenario without using PSO.

Table 12. REN PSO optimization result

Voyage 1			
	TRIPS	No-PSO (Freshwater)	PSO (Freshwater)
N1	JAKARTA-SAMARINDA	19.000	9.846
N2	SAMARINDA-BALIKPAPAN	0	8.881
N3	BALIKPAPAPAN-JAKARTA	0	0
Total Cost in Rupiah (in thousand scale)		1.235.000	817.000
Difference	33,85%		
Voyage 2			
N1	JAKARTA-SAMARINDA	23.359	9.846
N2	SAMARINDA-BALIKPAPAN	0	13.513
N3	BALIKPAPAPAN-SEMARANG	0	0
N4	SEMARANG-JAKARTA	0	0
Total Cost in Rupiah (in thousand scale)		1.508.000	910.000
Difference	40,40%		
Voyage 3			
N1	JAKARTA-BANJARMASIN	25.000	6.369
N2	BANJARMASIN-SAMARINDA	0	6.356
N3	SAMARINDA-SURABAYA	0	14.103
N4	SURABAYA-JAKARTA	1.829	1
Total Cost in Rupiah (in thousand scale)		1.689.105	918.540
Difference	45,63%		

In REN ship, the optimization effectively mitigates high initial costs at Jakarta (65,000/liters) by redistributing refills to lower-cost ports like Samarinda and Balikpapan. By aligning refills with route-specific consumption patterns, PSO achieves cost savings of 40.4% to 45.63%, demonstrating its effectiveness in adjusting refill quantities based on port price variability and consumption rates. To provide clearer results, the visualization results are given in the table and plot below.

Here in Table 13, presents 3 voyages from ASJ ship.

Table 13. AKA voyages

	AKA Voyages
1	SURABAYA-SAMARINDA;SAMARINDA-BOMBANA;BOMBANA-MAKASSAR;MAKASSAR-SURABAYA
2	MAKASSAR-BALIKPAPAN;BALIKPAPAN-SAMARINDA;SAMARINDA-BOMBANA;BOMBANA-SAMARINDA;SAMARINDA-MAKASSAR
3	SURABAYA-BALIKPAPAN;BALIKPAPAN-SAMARINDA;SAMARINDA-BATULICIN;BATULICIN-SURABAYA

Table 14 shows the cost results for the freshwater filling optimization for each voyage, along with a comparison to the scenario without using PSO.

Table 14. AKA PSO optimization result

	TRIPS	No-PSO (Freshwater filled)	PSO (Freshwater filled)
Voyage Sample 1			
N1	SURABAYA-SAMARINDA	27.000	7766
N2	SAMARINDA-BOMBANA	0	19.371
N3	BOMBANA-MAKASSAR	0	0
N4	MAKASSAR-SURABAYA	137	0
Total Cost in Rupiah (in thousand scale)		949.110	659.230
Difference	30,54%		
Voyage Sample 2			
N1	MAKASSAR-BALIKPAPAN	27.000	5593
N2	BALIKPAPAN-SAMARINDA	0	1023
N3	SAMARINDA-BOMBANA	0	8933
N4	BOMBANA-SAMARINDA	0	6321
N5	SAMARINDA-MAKASSAR	6.600	5460
Total Cost in Rupiah (in thousand scale)		810.000	669.832
Difference	17,31%		
Voyage Sample 3			
N1	SURABAYA-BALIKPAPAN	17.771	6578
N2	BALIKPAPAN-SAMARINDA	0	974
N3	SAMARINDA-BATULICIN	0	10.219
N4	BATULICIN-SURABAYA	0	0
Total Cost in Rupiah (in thousand scale)		621.895	457.986
Difference	26,37%		

PSO reduces costs by prioritizing lower-cost refills at Samarinda (20,000/liters) and strategically limiting purchases at more expensive ports. The strategy leverages lower-cost segments while ensuring sufficient supply for critical route sections, leading to cost savings ranging from 17.31% to 30.54% without compromising operational needs.

The PSO optimization effectively reduces freshwater refilling costs for ASJ, REN, and AKA vessels, achieving savings of 9.1%, 40.4%, and 30.5% respectively. The extent of cost reduction is influenced by the number of trips and price variability along the route, where more trips provide greater opportunities for optimization, and higher price variability allows PSO to strategically select lower-cost ports for refilling. Then to summarize and clarify how the results of XGBoost and PSO, the table below is provided.

Table 15. Summary of XGBoost and PSO results

		Freshwater consumption (liters)		Price result (rupiah, thousand scale)	
		No-XGBoost	XGBoost	No-PSO	PSO
ASJ	Voyage 1	15554	15549	1.181.660	1.074.030
	Voyage 2	6615	6511	1.196.660	1.079.672
	Voyage 3	2936	2966	1.763.540	1.558.205
REN	Voyage 1	8916	8973	1.235.000	817.000
	Voyage 2	5715	5694	1.508.000	910.000
	Voyage 3	8652	8593	1.689.105	918.540

AKA	Voyage 1	7774	7789	949.110	659.230
	Voyage 2	1542	1554	810.000	669.832
	Voyage 3	8478	8507	621.895	457.986

Table 15 summarizes the impact of integrating XGBoost and PSO across three representative cargo ships over multiple voyages. This is consistent with the discussion of the XGBoost results and their visualization in the scatter plot (Figure 3), where the XGBoost model produced freshwater consumption estimates that closely matched actual usage, with minor deviations typically within a few liters, demonstrating strong predictive reliability.

More significantly, the PSO model consistently resulted in substantial cost reductions across all voyages, with savings ranging from approximately 9% to 40% compared to conventional refill strategies.
These results only validate the combined framework's ability to support cost-effective and operationally accurate freshwater planning on cargo ship operations.

CONCLUSION

The XGBoost model demonstrated a consistent predictive performance across various configurations, with RMSE values ranging from 764.05 to 758.44 in 7 configuraions and a relative error of approximately 9.48%. As mentioned in XGBoost result, on average, the predicted freshwater consumption deviates by less than 10% from actual usage across various voyages. In the context of ship operations, this level of accuracy is **operationally acceptable**, particularly for medium-sized vessels like the feedermax to small panamax classes used in this study.

Given that ships often plan with conservative margins or buffer volumes for critical supplies like freshwater, this margin of error remains within a safe planning threshold. It enables shipping operators to make cost-sensitive port selection decisions without compromising the safety or sufficiency of freshwater supply during the voyage.

And the PSO model demonstrates its effectiveness in optimizing freshwater refilling costs by strategically allocating refilling points based on consumption patterns and port price variability, achieving cost savings of 9.1% (ASJ), 40.4% (REN), and 30.5% (AKA), resulting in great amounts of total cost reductions respectively. Its capability to dynamically adjust to varying freshwater prices and consumption demands makes it a robust tool for maritime logistics optimization.

In conclusion, the integration of XGBoost and PSO demonstrates the efficacy of combining predictive modeling and optimization techniques to effectively predict freshwater consumption and strategically reduce refilling costs in container vessel logistics. And not only demonstrates a practical solution for optimizing freshwater management in container vessel operations, enabling more accurate consumption prediction and significant cost reductions, but also contributes theoretically by showcasing the synergy between machine learning and metaheuristic optimization for solving complex, constraint-based planning problems in maritime logistics.

However, future research can explore not only more refined hyperparameter tuning and the integration of real-time port pricing data, but also the inclusion of additional operational variables such as weather patterns, voyage speed, and crew consumption behavior to improve prediction robustness. Moreover, expanding the model to support multi-ship or fleet level freshwater logistics, and incorporating adaptive optimization algorithms that respond to live data inputs, could significantly enhance its decision-making capabilities in dynamic maritime environments.

Overall, this study provides a concrete framework for data-driven resource planning in shipping logistics by uniting machine learning and metaheuristic optimization. The approach offers tangible cost-saving potential while laying the groundwork for intelligent maritime decision support systems, marking a meaningful contribution to both operational maritime efficiency and the advancement of applied computer science approach in real world industrial contexts.

REFERENCES

[1] E. K. Hansen, H. B. Rasmussen, and M. Lützen, "Making shipping more carbon-friendly? Exploring ship energy efficiency management plans in legislation and practice," *Energy Res Soc Sci*, vol. 65, p. 101459, Jul. 2020, doi: 10.1016/J.ERSS.2020.101459.

[2] X. Han *et al.*, "Water strategies and management: current paths to sustainable water use," *Appl Water Sci*, vol. 14, no. 7, pp. 1–14, Jul. 2024, doi: 10.1007/S13201-024-02214-2/FIGURES/1.

[3] J. Alferes *et al.*, "Water-smart strategies to support decision-making for water resource management in the industrial context," *Water Science and Technology*, vol. 90, no. 8, pp. 2276–2290, Oct. 2024, doi: 10.2166/WST.2024.326/1499537/WST2024326.PDF.

[4] J. Naseer, Z. B. Junaid, and A. Khatoon, "OPTIMIZING BLUE ECONOMY THROUGH EFFICIENT RESOURCE ALLOCATION: A DEEP DIVE INTO KARACHI SHIPYARD," *Journal of Research in Economics and Finance Management*, vol. 3, no. 1, pp. 53–71, Jun. 2024, doi: 10.56596/JREFM.V3I1.119.

[5] Z. Fei and Z. Li, "Coordinated Operation of A Green Multi-Energy Ship Microgrid with Hydrogen and Seawater Desalination," *IEEE Power and Energy Society General Meeting*, 2024, doi: 10.1109/PESGM51994.2024.10761086.

[6] J. Xing, J. Shen, Q. Pang, M. Fang, and H. Chen, "The finest diamond must be green: a closer look at the roles of institution in shipping firms' sustainable practices," *Environmental Science and Pollution Research*, vol. 30, no. 35, pp. 84631–84644, Jun. 2023, doi: 10.1007/S11356-023-28368-1/METRICS.

[7] Q. Zhou, S. Qu, Q. Wang, Y. She, Y. Yu, and J. Bi, "Sliding Window-Based Machine Learning for Environmental Inspection Resource Allocation," *Environ Sci Technol*, vol. 57, no. 44, pp. 16743–16754, Nov. 2023, doi: 10.1021/ACS.EST.3C05088/ASSET/IMAGES/LARGE/ES3C05088_0005.JPEG.

[8] S. Zheng, "Integration Strategies of Shipping Enterprises under Environmental Regulations from Perspective of Resource-Based View," *Advances in Economics, Management and Political Sciences*, vol. 139, no. 1, pp. 20–32, Dec. 2024, doi: 10.54254/2754-1169/2024.19200.

[9] C. Pan *et al.*, "An innovative process design of seawater desalination toward hydrogen liquefaction applied to a ship's engine: An economic analysis and intelligent data-driven learning study/optimization," *Desalination*, vol. 571, p. 117105, Feb. 2024, doi: 10.1016/J.DESAL.2023.117105.

[10] S. Asal, A. Acır, and I. Dincer, "A sustainable cruise ship development with cleaner production of electricity, heat, cooling, freshwater and hydrogen," *J Clean Prod*, vol. 467, p. 142939, Aug. 2024, doi: 10.1016/J.JCLEPRO.2024.142939.

[11] M. Bännstrand, A. Jönsson, H. Johnson, and R. Karlsson, "Study on the optimization of energy consumption as part of implementation of a ship energy efficiency management plan (SEEMP)," 2016.

[12] M. Ghasemi-Sardabrud, M. Zarkandi, and M. Mahmoodjanloo, "A Ship Routing and Scheduling Problem Considering Pickup and Delivery, Time Windows and Draft Limit," *Proceedings of 2019 15th Iran International Industrial Engineering Conference, IIIEC 2019*, pp. 165–170, May 2019, doi: 10.1109/IIIEC.2019.8720627.

[13] C. Sui, D. Stapersma, K. Visser, P. de Vos, and Y. Ding, "Energy effectiveness of ocean-going cargo ship under various operating conditions," *Ocean Engineering*, vol. 190, p. 106473, Oct. 2019, doi: 10.1016/J.OCEANENG.2019.106473.

[14] J. Sade, "Analysis Of The Placement And Needs Of General Cargo Ship Tanks With DWT 3650 Tons," *Maritime Park Journal of Maritime Technology and Society*, pp. 104–110, Oct. 2022, doi: 10.62012/MP.V1I3.21979.

[15] Y. Sang, Y. Ding, J. Xu, and C. Sui, "Ship voyage optimization based on fuel consumption under various operational conditions," *Fuel*, vol. 352, p. 129086, Nov. 2023, doi: 10.1016/J.FUEL.2023.129086.

[16] H. Yu, Z. Fang, X. Fu, J. Liu, and J. Chen, "Literature review on emission control-based ship voyage optimization," *Transp Res D Transp Environ*, vol. 93, p. 102768, Apr. 2021, doi: 10.1016/J.TRD.2021.102768.

[17] M. G. Borg, C. D. Muscat-Fenech, T. Tezdogan, T. Sant, S. Mizzi, and Y. K. Demirel, "A Numerical Analysis of Dynamic Slosh Dampening Utilising Perforated Partitions in Partially-Filled Rectangular Tanks," *Journal of Marine Science and Engineering 2022, Vol. 10, Page 254*, vol. 10, no. 2, p. 254, Feb. 2022, doi: 10.3390/JMSE10020254.

[18] C. Zhang, X. Zou, and C. Lin, “Fusing XGBoost and SHAP Models for Maritime Accident Prediction and Causality Interpretability Analysis,” *Journal of Marine Science and Engineering 2022, Vol. 10, Page 1154*, vol. 10, no. 8, p. 1154, Aug. 2022, doi: 10.3390/JMSE10081154.
[19] P. Han, Z. Liu, Z. Sun, and C. Yan, “A novel prediction model for ship fuel consumption considering shipping data privacy: An XGBoost-IGWO-LSTM-based personalized federated learning approach,” *Ocean Engineering*, vol. 302, p. 117668, Jun. 2024, doi: 10.1016/J.OCEANENG.2024.117668.
[20] V. N. Nguyen, N. Chung, G. N. Balaji, K. Rudzki, and A. T. Hoang, “Internet of things-driven approach integrated with explainable machine learning models for ship fuel consumption prediction,” *Alexandria Engineering Journal*, vol. 118, pp. 664–680, Apr. 2025, doi: 10.1016/J.AEJ.2025.01.067.
[21] H. Hu, A. J. van der Westhuysen, P. Chu, and A. Fujisaki-Manome, “Predicting Lake Erie wave heights and periods using XGBoost and LSTM,” *Ocean Model (Oxf)*, vol. 164, p. 101832, Aug. 2021, doi: 10.1016/J.OCEMOD.2021.101832.
[22] A. S. Sajjanshetty, V. Jayanth, R. Mohan, S. Pahari, and C. Deepti, “Estimation of Community Water Consumption Using Multivariate Ensemble Approach.,” *Proceedings of IEEE InC4 2023 - 2023 IEEE International Conference on Contemporary Computing and Communications*, 2023, doi: 10.1109/INC457730.2023.10263265.
[23] M. Jahandideh-Tehrani, O. Bozorg-Haddad, and H. A. Loáiciga, “Application of particle swarm optimization to water management: an introduction and overview,” *Environ Monit Assess*, vol. 192, no. 5, p. 281, May 2020, doi: 10.1007/S10661-020-8228-Z.
[24] A. Waluyo, H. Jatnika, M. R. S. Permatasari, T. Tuslaela, I. Purnamasari, and A. P. Windarto, “Data Mining Optimization uses C4.5 Classification and Particle Swarm Optimization (PSO) in the location selection of Student Boardinghouses,” *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, Jul. 2020, doi: 10.1088/1757-899X/874/1/012024.
[25] B. Lu, M. Zhang, Z. Yan, and G. Yu, “Optimization of pump allocation in real water networks for water quality improvement using particle swarm optimization,” *OCEANS 2023 - Limerick, OCEANS Limerick 2023*, 2023, doi: 10.1109/OCEANSLIMERICK52467.2023.10244548.
[26] M. Fuentetaja-Merino, A. Silva-Campillo, M. A. Herreros-Sierra, and F. Pérez-Arribas, “Structural Analysis of the Geometric Alternatives of Double-Bottom Floor Plates of a Panamax-Class Container Ship,” *Applied Sciences 2024, Vol. 14, Page 10684*, vol. 14, no. 22, p. 10684, Nov. 2024, doi: 10.3390/APP142210684.
[27] B. Jeong and K. Kim, “Characteristics of Economic and Environmental Benefits of Shore Power Use by Container-Ship Size,” *Journal of Marine Science and Engineering 2022, Vol. 10, Page 622*, vol. 10, no. 5, p. 622, May 2022, doi: 10.3390/JMSE10050622.
[28] O. Polat and O. Polat, “Designing of Container Feeder Service Networks Under Unstable Demand Conditions,” *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-8709-6.ch006*, pp. 115–136, Jan. 1AD, doi: 10.4018/978-1-7998-8709-6.CH006.
[29] B. Jeong and K. Kim, “Characteristics of Economic and Environmental Benefits of Shore Power Use by Container-Ship Size,” *Journal of Marine Science and Engineering 2022, Vol. 10, Page 622*, vol. 10, no. 5, p. 622, May 2022, doi: 10.3390/JMSE10050622.
[30] A. Ben Hamida, M. Kacem, C. de Peretti, and L. Belkacem, “Machine learning based methods for ratemaking health care insurance,” *International Journal of Market Research*, Nov. 2024, doi: 10.1177/14707853241275446;JOURNAL:JOURNAL:MREA;PAGE:STRING:ARTICLE/CHAPTER.
[31] H. T. Wen, H. Y. Wu, K. C. Liao, and W. C. Chen, “JT9D Engine Thrust Estimation and Model Sensitivity Analysis Using Gradient Boosting Regression Method,” *Aerospace 2023, Vol. 10, Page 639*, vol. 10, no. 7, p. 639, Jul. 2023, doi: 10.3390/AEROSPACE10070639.
[32] J. Wang and S. Zhou, “CS-GA-XGBoost-Based Model for a Radio-Frequency Power Amplifier under Different Temperatures,” *Micromachines 2023, Vol. 14, Page 1673*, vol. 14, no. 9, p. 1673, Aug. 2023, doi: 10.3390/MI14091673.
[33] J. Tian *et al.*, “Synergetic Focal Loss for Imbalanced Classification in Federated XGBoost,” *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 2, pp. 647–660, Feb. 2024, doi: 10.1109/TAI.2023.3254519.
[34] B. Shi, C. Tan, and Y. Yu, “Predicting the S&P 500 stock market with machine learning models,” *Applied and Computational Engineering*, vol. 48, no. 1, pp. 255–261, Mar. 2024, doi: 10.54254/2755-2721/48/20241621.

[35] J. R. Vage, S. Ghose, R. P. Deb Nath, D. Ghose, Y. Lin, and S. P. Dash, “Predictive Modeling for Heart Rate: A Comparative Analysis of LSTM, XGBoost, and LightGBM,” *2023 26th International Conference on Computer and Information Technology, ICCIT 2023*, 2023, doi: 10.1109/ICCIT60459.2023.10441516.

[36] A. Demir Yetiş, N. İlhan, and H. Kara, “Integrating deep learning and regression models for accurate prediction of groundwater fluoride contamination in old city in Bitlis province, Eastern Anatolia Region, Türkiye,” *Environmental Science and Pollution Research*, vol. 31, no. 34, pp. 47201–47219, Jul. 2024, doi: 10.1007/S11356-024-34194-W/TABLES/6.

[37] F. H. Rizk, A. Saleh, A. Elgaml, A. Elsakaan, and A. M. Zaki, “Exploring Predictive Models for Students’ Performance in Exams: A Comparative Analysis of Regression Algorithms,” *Journal of Artificial Intelligence and Metaheuristics*, vol. 7, no. 1, pp. 38–52, 2024, doi: 10.54216/JAIM.070103.

[38] U. Umoh, D. Asuquo, I. Eyoh, A. Abayomi, E. Nyoho, and H. Vincent, “A Fuzzy-Based Support Vector Regression Framework for Crop Yield Prediction,” pp. 173–185, 2022, doi: 10.1007/978-981-16-1740-9_16.

[39] P. Dumre, S. Bhattarai, and H. K. Shashikala, “Optimizing Linear Regression Models: A Comparative Study of Error Metrics,” *Proceedings - 4th International Conference on Technological Advancements in Computational Sciences, ICTACS 2024*, pp. 1856–1861, 2024, doi: 10.1109/ICTACS62700.2024.10840719.

[40] D. P. Garapati, L. V. A. P. Maddipati, A. Dedeepya, M. M. Jonnalagadda, J. L. V. Kandulapati, and S. N. Gundu, “Comparative Analysis of Regression Algorithms in Solar Power Production Forecasting,” *2nd IEEE International Conference on Networks, Multimedia and Information Technology, NMITCON 2024*, 2024, doi: 10.1109/NMITCON62075.2024.10698849.