POLISH MARITIME RESEARCH 2 (122) 2024 Vol. 31; pp. 140-155
10.2478/pomr-2024-0030

Leveraging Artificial Intelligence to Enhance Port Operation Efficiency

Gia Huy Dinh [1]
Hoang Thai Pham [1]
Lam Canh Nguyen [2]
Hai Quoc Dang [1]
Nguyen Dang Khoa Pham [1, 4 *]
[1] University of Transport Ho Chi Minh City, Viet Nam
[2] RMIT University, Viet Nam
[4] PATET Research Group, Viet Nam

* Corresponding author: *khoapnd@ut.edu.vn (Nguyen Dang Khoa Pham)*

ABSTRACT

Maritime transport forms the backbone of international logistics, as it allows for the transfer of bulk and long-haul products. The sophisticated planning required for this form of transportation frequently involves challenges such as unpredictable weather, diverse types of cargo kinds, and changes in port conditions, all of which can raise operational expenses. As a result, the accurate projection of a ship's total time spent in port, and the anticipation of potential delays, have become critical for effective port activity planning and management. In this work, we aim to develop a port management system based on enhanced prediction and classification algorithms that are capable of precisely forecasting the lengths of ship stays and delays. On both the training and testing datasets, the XGBoost model was found to consistently outperform the alternative approaches in terms of RMSE, MAE, and R2 values for both the turnaround time and waiting period models. When used in the turnaround time model, the XGBoost model had the lowest RMSE of 1.29 during training and 0.5019 during testing, and also achieved the lowest MAE of 0.802 for training and 0.391 for testing. It also had the highest R2 values of 0.9788 during training and 0.9933 during testing. Similarly, in the waiting period model, the XGBoost model outperformed the random forest and decision tree models, with the lowest RMSE, MAE, and greatest R2 values in both the training and testing phases.

Keywords: Port management; Artificial intelligence; Machine learning; Sustainable maritime; Logistics efficiency.

INTRODUCTION

In the intricate and ever-changing field of maritime operations, which includes port operation and management, ship and vessel traffic management, logistics, shipbuilding, and security, the demand for creative solutions to improve efficiency and gain a competitive edge is at its peak [1–4]. Due to its ability to process large amounts of data, predict patterns, and automate decision-making processes, artificial intelligence (AI) is revolutionising port operations [5–7]. An exploration of the various applications of AI illustrates the advantages and the obstacles encountered when incorporating it into port operations, which include the integration of technologies such as IoT and blockchain and the refinement of operational procedures to boost efficiency, cut down on turnaround times, and optimise resource usage to lower costs and reduce greenhouse gas (GHG) emissions [8–12]. The use of AI to enhance cargo handling processes can greatly accelerate loading and unloading activities, as shown by its implementation at the Port of Rotterdam for predicting the best container placement on ships [13,14]. In addition, the use of AI technology for predictive maintenance enables the monitoring of equipment sensors to anticipate failures in advance, as demonstrated at the Port of Los Angeles, resulting in decreased operational interruptions and prolonged machinery/engine lifespans [15–18].

AI can help in improving port security and surveillance when automated systems are used to monitor CCTV feeds in real time [19], as it can help in detecting abnormal behavior or unauthorised entries, thus strengthening the security measures of ports [20]. However, despite these benefits, there are certain obstacles to overcome, such as the upfront costs of implementing AI systems, the requirement for ongoing data input for improvements to the system, and the importance of training employees to work effectively with advanced technology [21,22]. Smart technologies can help in effectively managing energy use in port operations, resulting in notable decreases in energy consumption [23–27], and the use of automated guided vehicles and electric cranes can reduce dependence on non-renewable energy sources, resulting in cost savings and decreased carbon emissions [28,29]. Through the digitisation of processes and documentation, operations can become more efficient by reducing ship idle times in port; this results in decreased fuel consumption and GHG emissions, ultimately lowering port fees and fuel costs for shipping companies [30–32]. Improving port management by enhancing planning and forecasting helps alleviate congestion and streamline ship schedules, ultimately cutting down on idle time for ships and lowering fuel usage and emissions [33,34].

As reported in the literature, optimising the fuel consumption for ship operation [35–37] through the use of green and alternative fuels for ship or internal combustion engine-based equipment in a port [38–41] or through the electrification of port equipment and ships [42–44] can also offer the potential to achieve low GHG and pollutant emissions. By strategically integrating technology and enhancing operational practices, ports can lower expenses, minimise their environmental impact, and support global efforts to decrease GHG emissions and advance environmental sustainability in the maritime industry [45–47]. Modern technologies, including AI, IoT and blockchain technology, have altered the way companies operate in the modern age. AI, IoT and blockchain technology can be used for many applications [48,49]. According to the findings of a study by Xu et al. [50], the adoption of AI technology by various ports has the potential to increase port profitability; however, the unfortunate reality is that the simultaneous adoption of this technology makes homogenised competition even more intense, which poses a risk to the realisation of profits. In addition, though a hub port has the ability to harness the benefits of AI to increase its competitiveness, it also has the potential to harm the performance of rivals and society as a whole.

A holistic strategy that involves upgrading port and logistics operations with AI and other cutting-edge technologies, despite obstacles related to integration, will give rise to a future where ports function with unparalleled accuracy and productivity [51,52]. These innovations offer numerous advantages, such as improving the sustainability of port operations and advancing the management of activities within ports, which promises to boost global trade efficiency and promote environmental stewardship [53–55]. Maritime transport forms the foundation of international logistics, as it permits the delivery of bulk and long-haul items. However, the intricate planning necessary for marine transportation frequently involves challenges such as unexpected weather, a wide range of types of cargo, and changes in port conditions. These problems can have a substantial influence on operating costs, meaning that it is critical to precisely assess a ship's entire stay in port and to anticipate unexpected delays. Thus, in this work, we focus on addressing the inherent constraints of marine logistics through the use of AI to improve port operations.

METHODOLOGY

PROBLEM STATEMENT

The problem of berth allocation is a critical one in port administration, as it requires the allocation of available berths to arriving ships in a way that improves efficiency while minimising delays. An example will illustrate the difficulty of berth allocation: consider a crowded container terminal at a seaport that receives many cargo ships of varying sizes and with differing arrival times during the day. The terminal has a limited number of berths where ships may dock to load and unload cargo. The aim is to best distribute berths to arriving ships to ensure smooth operation and reduce waiting periods.

In this scenario, machine learning (ML) is used to anticipate the waiting and turnaround times of ships in port, in order to improve operational efficiency at a container terminal with four berths, each capable of housing one container ship at a time. Throughout the day, several ships arrive with their own particular characteristics, such as arrival time, expected loading/unloading time (turnaround time), and priority level, which are determined based on the cargo type, size of the vessel, and contractual agreements. The task entails improving numerous aspects to enhance port operations based on how ships are controlled depending on their arrival times to avoid excessive delays. ML models are used to precisely anticipate ship turnaround times, thus enabling more effective scheduling and resource usage. The objective is to create a predictive model that can accurately estimate ship waiting and turnaround times, which can contribute to the management of port operations. This involves the use of advanced ML algorithms that assess real-time data on ship arrivals and operational restrictions, allowing for more informed decision-making.

The present study is innovative in that we apply a comprehensive approach to enhancing port efficiency by integrating ML predictions with operational decision-making. Unlike previous techniques, which mainly rely on heuristic or rule-based systems to manage ship arrivals and resource allocation, we use advanced predictive models to precisely estimate ship waiting periods and turnaround times. This predictive capacity enables proactive revisions to operational plans, resulting in significantly reduced ship idle periods and better usage of port resources. Furthermore, our method takes into account a wide variety of factors influencing port operations, from ship-specific features to larger operational restrictions, resulting in a more nuanced and successful strategy for port management. The use of ML models that are capable of learning from previous data and reacting to new information distinguishes this study, making it an important contribution to the field of port operation improvement. The proposed technique not only advances the theoretical understanding of the use of ML to solve logistical difficulties, but also provides a practical framework that may be beneficial to ports around the globe in terms of improving efficiency, lowering costs, and improving service quality.

MACHINE LEARNING

ML can help solve the berth allocation problem by combining historical data with real-time information and optimisation algorithms to make data-driven judgments. In the following, we describe two ML approaches that can be employed to tackle the problem.

Random Forest

A random forest (RF) algorithm consists of an ensemble of N decision trees, $\{T_1, T_2, \ldots, T_N\}$. Each tree T_i is built from a bootstrap sample of the training data, D, which is a sample drawn with replacement from the original training dataset of size M. This method is also sometimes referred to as bagging or bootstrap aggregating. In addition, when a node is split during the building of the tree, rather than searching for the most optimal division among all features, a random subset of k features is selected from the total K features, and the optimum split from this subset is used to divide the node. This is done in place of searching for the best split among all features. The presence of this unpredictability contributes to the model's increased robustness and helps to prevent overfitting [56–59].

Training process: Consider a dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$ where x_i is a vector of features and y_i is the objective variable for each instance i. The objective of the RF regressor is to learn a model that can predict the value of y for new instances based on their features x.

For each tree T_i in a RF model F, the training process involves:

- Generation of the bootstrap sample D_i for the original dataset D.
- Recursively splitting D_i at each node, beginning with the root, until the stopping criteria are met (e.g., maximum depth, minimum samples at a leaf), using the best split chosen from a randomly selected subset of k features at each step.

Prediction process: The forecast for a new instance x is derived by taking the average of the predictions provided by each of the distinct trees in the forest:

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x) \tag{1}$$

where $T_i(x)$ denotes the forecast by tree for x, and the total number of trees is denoted as N.

RF is highly effective in minimising prediction errors by reducing the variance without substantially increasing the bias. If we assume that the trees are uncorrelated and each tree's prediction has a variance of σ^2, the mean prediction variance from the RF algorithm is . It can be seen that an increase in the number of trees N decreases the prediction variance. An ensemble learning method combines predictions from multiple ML algorithms (in this case, decision trees) to improve the accuracy, and a typical flow chart for this process is depicted in **Fig. 1**. Through the integration of various predictions from multiple trees, RF can offer precise results, process extensive datasets with increased dimensionality, and address the problem of missing values efficiently [60–62].

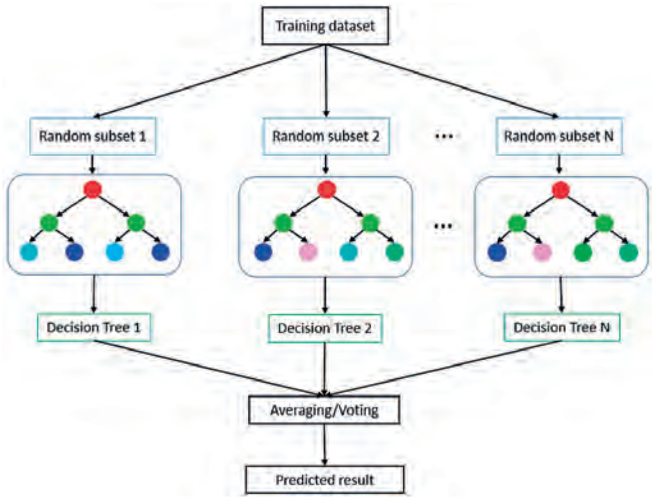


Fig. 1. Flow chart for random forest regression [63]

eXtreme Gradient Boosting

The gradient boosting framework has gained popularity due to its speed and performance in ML contests. eXtreme gradient boosting (XGBoost) is an enhanced and fast implementation of the gradient boosting framework. However, in contrast to RF, which constructs and aggregates many decision trees in parallel with no interaction, XGBoost constructs trees in a sequential manner, with each new tree rectifying the faults of trees that were constructed in the past. The use of this strategy makes it possible to create a model that is more optimised and is capable of handling difficult regression and classification tasks with a high level of efficiency. XGBoost is a methodical technique for eliminating mistakes and enhancing model performance, and the mathematical expressions and concepts that underpin it illustrate this approach [64,65]. There are two components that make up the objective function of XGBoost, which aims to minimise the loss function. These components are the loss function and the regularisation term. In order to prevent overfitting, the regularisation term is responsible for controlling the complexity of the model, while the loss function is responsible for evaluating the difference between the predicted values and the actual values. A schematic diagram of the operation of XGBoost is shown in **Fig. 2**.

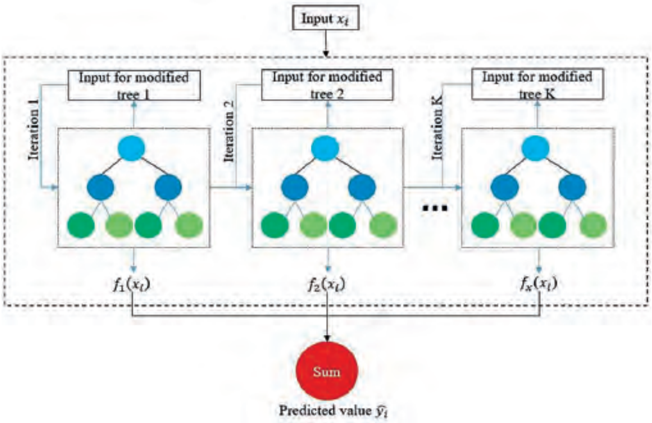


Fig. 2. Schematic diagram of XGBoost [66]

Given a dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$, in the case when x_i is the feature vector of the i-th instance and y_i is its corresponding target value, the objective function at iteration t can be expressed as:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega\left(f_t\right) \quad (2)$$

Where $(y_i, \hat{y}_i^{(t-1)}$ is the forecast of the i-th instance at the (t–1)-th iteration, while $f_t(x_i)$ is the forecast of the t-th tree and l is the differentiable convex loss function, which measures the difference between the forecast and observed values. The term $\Omega(f_t)$ is the regularisation term for the t-th tree, which often reflects both the intricacy of the tree (for example, the number of leaves) and the L2 norm of the leaf weights [67,68].

Training process: XGBoost's training technique involves adding trees in an iterative manner, where each tree is taught to correct the residuals (or mistakes) of the previous ensemble of trees. The model is modified by introducing a function ft that optimises the goal function. This entails calculating the gradient and Hessian of the loss function for the predictions and using these to select the direction in which the tree should grow.

Forecasting process: The forecast for a new instance (x) after T rounds of boosting is the sum of the predictions from all T trees:

$$\hat{y}(x) = \frac{1}{N} \sum_{i=1}^{T} f_i(x) \quad (3)$$

Optimisation techniques: XGBoost can be used with numerous optimisation methods to improve the efficiency and speed, for example:

- Gradient-based one-side sampling (GOSS): This reduces the data size for faster performance while maintaining algorithm efficacy.
- Regularization: To avoid overfitting, the objective function contains both L1 (lasso regression) and L2 (ridge regression) regularisation terms.
- Sparsity-aware split finding: This approach efficiently manages missing data by either discovering the optimal way to handle missing values during training or assigning them a specific value.

Decision Tree Regressor

The decision tree (DT) regressor is a simple and effective ML technique used for regression applications. It works by recursively splitting the feature space into different areas and estimating the average target value of the training cases within each zone. DTs are frequently used in a wide variety of fields due to their simplicity and interpretability [69–71].

Foundation of DT: A DT is a hierarchical structure of nodes, where each node represents a feature, and each edge represents a decision rule based on that feature. The goal of a DT algorithm is to recursively divide the feature space into sections that are as homogenous as possible in terms of the target variable. Given a dataset D = {$(x_1, y_1),(x_2, y_2),\ldots,(x_M, y_M)$}, where x_i-th is the feature vector for the i-th instance and y_i is the target objective, a typical schematic diagram for a DT algorithm is shown in **Fig. 3**.

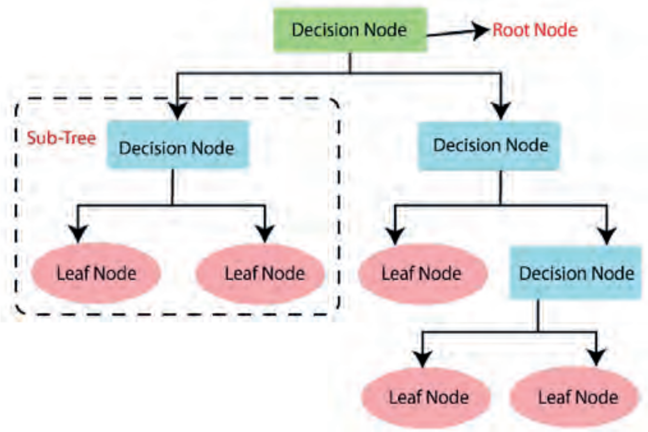


Fig. 3. Schematic diagram of a decision tree algorithm [72]

Model training: A DT model is trained by recursively dividing the feature space into regions, with the goal of minimising the variance of the target variable across each zone. The algorithm starts at the root node and passes through each node, selecting the optimal split based on a splitting criterion such as the mean squared error (MSE) or the mean absolute error (MAE). The splitting criterion is used to assess the quality of a split by calculating the reduction in variance or absolute error that occurs. The procedure continues until a stopping requirement is satisfied, such as achieving a maximum depth, sampling a certain number of leaves, or reducing impurities to a minimum [73–75].

From a numerical perspective, this splitting can be expressed as

$$J(s,t) = \frac{m_L}{m} Var(y_L) + \frac{m_R}{m} Var(y_R) \quad (4)$$

where s is the feature and threshold used for splitting, t is the threshold value, mL and mR are the number of instances in the left and right child nodes, m is the total number of instances, and Var(yL) and Var(yR) are the variances in the target values for the left and right child nodes, respectively.

The prediction for a new instance x is formed by traversing the DT from root to leaf node and returning the average target value of the training examples within the leaf node. The mathematical expression for the prediction is as follows:

$$\hat{y}(x) = \frac{1}{m_x} \sum_{i=1}^{m_x} y_i \quad (5)$$

where m_x corresponds to the number of instances in the leaf node containing x and y, which is the target value of the i-th instance.DT regression provides a simple but efficient approach for forecasting and modeling continuum target variables. DT makes regression tasks more visible and understandable by recursively splitting the feature space and making predictions based on the average target values inside each zone [76]. Despite certain shortcomings, DTs are still a popular option for regression issues due to their simplicity, adaptability, and ease of understanding.

MODEL EVALUATION TECHNIQUES

The coefficient of determination (R^2), root mean squared error (RMSE), and MAE are key metrics for assessing the performance of a regression model. They give information about how well the model matches the data and the accuracy of its predictions. The mathematical expressions and principles that underpin these measures explain their relevance in measuring the quality of regression models [77,78].

Coefficient of determination (R^2): This estimates the amount of variation in the dependent variable that is predicted from the independent variables. Its value ranges from zero to one, with a higher value indicating a better match between the model and the data [79–82]where the coefficient of determination (R2.

R^2 may be stated mathematically as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (6)$$

where SS_{res} is the sum of squared residuals (the difference between the actual and predicted values) and SS_{tot} is the total sum of squares (the variance in the dependent variable).

RMSE: This represents the average magnitude of the errors between the actual and anticipated values. It gives an idea of the typical deviation of the forecasts from the actual values, with smaller values indicating better model performance. The RMSE may be mathematically represented as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (7)$$

where n is the number of instances, y_i is the actual value of the objective variable of the i-th instance, and $\hat{y}_i$ is the forecast value.

MAE: This is similar to RMSE, except it quantifies the average absolute variation between the forecasts and actual values. Like RMSE, lower MAE values imply higher model performance.

The MAE may be mathematically stated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (8)$$

A higher R^2 value indicates that the model explains more of the variation in the dependent variable. Its value ranges from zero to one, with one representing a perfect match. The RMSE and MAE are two metrics that assess prediction accuracy, with lower values indicating better performance. However, RMSE penalises large mistakes more severely than MAE, making it more susceptible to outliers [83–85]including the oil and gas industry, which covers several fields, including reservoirs, drilling, and production. In oil and gas production, conventional methods, such as reservoir simulation, are used to predict the oil production rate. This simulation requires comprehensive data, so each process step takes a long time and is expensive. AI is urgently needed and can be a solution in this case. This research aims to apply AI techniques to forecast oil production rates based on water injection rates from two injection wells. Three wells are connected with a direct line drive pattern. Three different AI methods were applied, including multiple linear polynomial regression (PR.

Evaluation of a model using R^2, RMSE, and MAE is a typical method for comparing regression models. These metrics give information about how well the model matches the data and the accuracy of its predictions.

Model selection: These measures can be used to select models with greater R^2 values, and lower RMSE and MAE values. In summary, R^2, RMSE, and MAE are critical metrics for evaluating the validity of regression models, as they give useful insights into the models' accuracy and goodness of fit, which aids in model evaluation, selection, and enhancement [58,86].

RESULTS AND DISCUSSION

CASE STUDY

Located in Vietnam, Haiphong Port plays a crucial role as a maritime hub that supports trade and commerce. The efficient management of ship waiting periods and turnaround times is crucial for optimising port operations and ensuring a smooth flow of cargo. In this case study, the goal is to create predictive models using advanced ML techniques such as

XGBoost, DT, and RF. These models aim to predict ship port waiting times and turnaround durations by exploiting the vast amount of data from the automatic identification system (AIS). The AIS provides up-to-date information on various factors related to vessel movements, such as locations, speeds, and headings. Using this extensive dataset, our goal is to create models that can predict waiting times and turnaround times for ships at Haiphong Port. These forecasts are crucial to enable port authorities to coordinate operations, optimise resource allocation, and improve overall port efficiency [87–89].

XGBoost, DT, and RF are well-regarded ML algorithms that are recognised for their effectiveness in addressing regression tasks. Through the use of these algorithms, we aim to leverage the complex patterns in the AIS data to produce precise forecasts. With a deep understanding of the factors affecting waiting periods and turnaround times, port authorities can proactively handle vessel traffic, alleviate congestion, and guarantee prompt processing of cargo. The creation of predictive models based on AIS data and advanced ML techniques represents an important development for Haiphong Port, as data-driven insights can allow port authorities to effectively address operational challenges, and to achieve enhanced efficiency and reliability in maritime logistics. With this in mind, we aim to streamline operations, decrease turnaround times, and enhance the competitiveness of Haiphong Port in terms of global trade.

Data Analysis

Preparing the data is a crucial stage when creating ML models, in order to guarantee precise and dependable forecasts. Prior to inputting data into ML algorithms, it is essential to thoroughly clean, transform, and organise the data appropriately, which improves the data quality and boosts the performance of the ML model. The correlation matrix in **Table 1** shows the connections between various ship and berth dimensions, turnaround times, and waiting time variables.

Tab. 1. Correlational matrix of the data

	Ship_L, m	**Ship_D, m**	**Berth_L, m**	**Berth_D, m**	**Turnaround_T, hr**	**Waiting_T, hr**
Ship_L, m	1	−0.122	−0.018	−0.108	−0.25	−0.102
Ship_D, m	−0.123	1	−0.14258	0.0997	0.11	0.067
Berth_L, m	−0.018	−0.143	1	0.116	−0.23	−0.184
Berth_D, m	−0.108	0.0997	0.115	1	0.0158	−0.054
Turnaround_T, hr	−0.25	0.11	−0.232	0.0158	1	0.0061
Waiting_T, hr	−0.10	0.067	−0.1843	−0.0541	0.0061	1

Each cell of the matrix displays the correlation coefficient between two variables. The ship length (Ship_L) has a negative correlation with turnaround time (−0.25) and waiting time (−0.102), suggesting that longer ships typically experience shorter turnaround and waiting times, although these correlation coefficients have a relatively low strength. Ship draft (Ship_D) has a slight positive correlation with turnaround time (0.11) and waiting time (0.067), indicating that a rise in ship draft could lead to a slight increase in both turnaround and waiting times. Berth length (Berth_L) shows weak negative correlations with both turnaround time (−0.23) and waiting time (−0.184), suggesting that ships with extended berths have reduced turnaround and waiting times. In contrast, berth draft (Berth_D) has relatively low correlation coefficients with other variables, suggesting weak correlations [90,91].

The turnaround time (Turnaround_T) has moderately negative correlations with ship length (−0.25) and berth length (−0.23), suggesting that longer ships with longer berths are linked to shorter turnaround times. Conversely, waiting time (Waiting_T) exhibits minimal correlations with ship and berth dimensions, with coefficients of close to zero. There are very minor negative correlations between ship length (−0.102) and berth length (−0.184). These correlation coefficients offer valuable insights into the connections among various variables, which are crucial for improving port operations and increasing efficiency. Through data preprocessing and identifying these relationships, we can create more precise ML models to forecast ship waiting times and turnaround durations, which will ultimately streamline port activities and improve effectiveness.

An examination of the statistical data offers interesting insights into the features and spread of the information, as shown in **Table 2**, which presents a summary of statistics for the variables related to ship and berth dimensions, turnaround time, and waiting time. The average ship length is 154.72 m, with a standard deviation of 28.5 m. Ship lengths range from 92.08 to 201.27 m, with ship lengths at the 25th percentile, median, and 75th percentile being 134.8, 160.98, and 173.87 m, respectively [92,93]rainfall, and humidity indirectly affect DHF spread patterns. Therefore, this research uses and compares three machine learning modelsâ€”restricted Boltzmann machine-backpropagation neural network (RBM-BPNN.

Tab. 2. Descriptive statistical analysis of the data

	Ship_L, m	Ship_D, m	Berth_L, m	Berth_D, m	Turnaround_T, hr	Waiting_T, hr
Mean	154.72	10.69	232.13	15.77	23.03	5.92
Std	28.5	2.15	30.25	2.04	8.47	1.84
Min	92.08	7.36	174.92	12.35	4.92	2.57
25%	134.8	9.06	213.45	15.65	16.97	4.58
50%	160.98	10.67	233.87	15.88	21.65	5.64
75%	173.87	12.29	259.55	17.99	28.65	7.24
Max	201.27	15.08	276.44	17.99	42.28	11.57

The average ship draft is 10.69 m, with a standard deviation of 2.15 m, and the ship drafts range from 7.36 to 15.08 m. The ship drafts at the 25th percentile, median, and 75th percentile are 9.06, 10.67, and 12.29 m, respectively.

The average berth length is 232.13 m, with a deviation of 30.25 m. The berth lengths range from 174.92 to 276.44 m, with the lengths of the 25th percentile, median, and 75th percentile berths being 213.45, 233.87, and 259.55 m, respectively.

The average berth draft is 15.77 m, with a standard deviation of 2.04 m. The berth drafts range from 12.35 to 17.99 m. The 25th percentile, median, and 75th percentile berth drafts are 15.65, 15.88, and 17.99 m, respectively.

The average turnaround time is 23.03 h, with a standard deviation of 8.47 h. The shortest and longest turnaround times are 4.92 and 42.28 h, respectively. The turnaround times for the 25th percentile, median, and 75th percentile are 16.97, 21.65, and 28.65 h, respectively.

The mean waiting time is 5.92 h, with a standard deviation of 1.84 h. The waiting times range from 2.57 to 11.57 h. The waiting times at the 25th percentile, median, and 75th percentile are 4.58, 5.64, and 7.24 h, respectively.

These statistics offer valuable insights into the distribution and characteristics of the data, which are important in terms of understanding the dataset and guiding subsequent analysis and modeling.

Data division: For model development and evaluation, the dataset is split into three subsets, to create the training, validation, and test sets. The training set makes up 70% of the dataset and forms the basis for training the ML models. This subset contains sufficient data to allow the models to grasp the fundamental patterns and connections in the data. Following this, 15% of the dataset is set aside for the validation set, which is used to assess the model's performance and to adjust the hyperparameters. This dataset is important for fine-tuning the model's setup to reach peak performance. By making iterative adjustments that are informed by the validation results, models can be fine-tuned to perform well on new data. Finally, 15% of the dataset forms the test set, which allows us to independently evaluate the model's performance. By isolating this subset from the training and validation data, we can guarantee a fair assessment of the model's capacity to adapt to fresh, unobserved data.

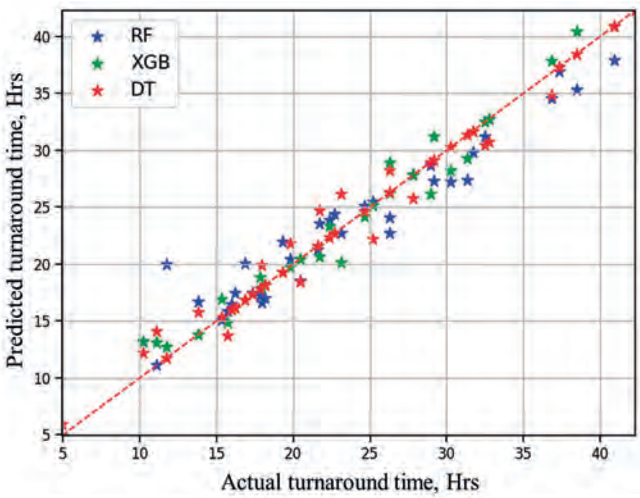
Hyperparameter optimisation: Bayesian optimisation is a robust method for effectively navigating the hyperparameter space and identifying the best configuration for each model. Through strategic sampling of the hyperparameter space based on previous evaluations, Bayesian optimisation speeds up the quest for the optimal model configuration. To provide guidance for the optimisation process, an objective function is established that involves minimising either the MAE or RMSE on the validation set. This function acts as a key metric enabling us to assess the performance of various hyperparameter configurations and to identify the most promising ones [94,95].

K-cross fold: Integrating a five-fold cross-validation strategy into the Bayesian optimisation process helps prevent overfitting and ensures the robustness of the models. Through the process of dividing the data into five subsets and repeatedly training and validating the models on various combinations of these subsets, we can achieve more dependable assessments of model performance and hyperparameter efficacy [96,97].

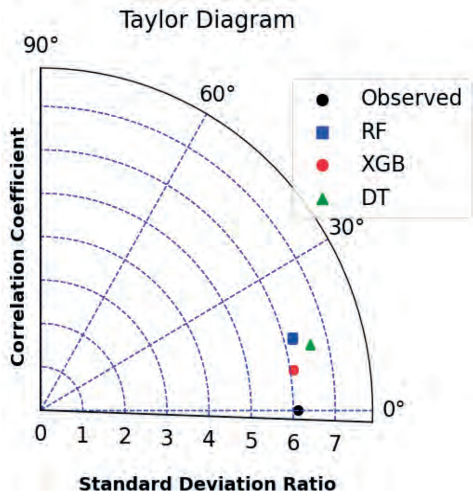
Turnaround Time Model

The XGBoost regressor was trained on the training set, with hyperparameters selected through Bayesian optimisation. XGBoost's robust gradient boosting framework allows for effective understanding of intricate patterns in the data, resulting in top-performing predictive models. In the same way, a DT regressor was trained on the training set using the default hyperparameters. DT provides a straightforward and easy-to-understand approach, which is useful for establishing a foundation model for comparison against more intricate algorithms. Following this, the RF regressor was trained with hyperparameters obtained through Bayesian optimisation using the same training set. Through the combination of various individual decision trees, RF addresses the problem of overfitting and gives enhanced predictive accuracy, and is a valuable asset in the modeling ensemble [98,99].

Once the prediction models had been developed, they were tested for predictions on the data. The model output during the training and testing phases is depicted in **Figs. 4a** and **5a**. A comparison of the results reveals that the XGBoost model performs the best of the three models, followed by DT and then RF. The models were also compared using Taylor diagrams, as shown in **Figs. 4b** and **5b**. It was easiest to compare the models using Taylor diagrams, as it is straightforward to identify the best-performing models using this approach. It can be observed that XGBoost-based models gave the best performance of the three considered here. Some statistical values are given in **Table 3**.

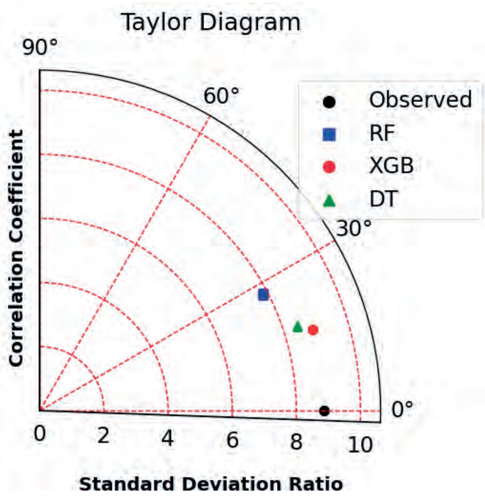


(a)

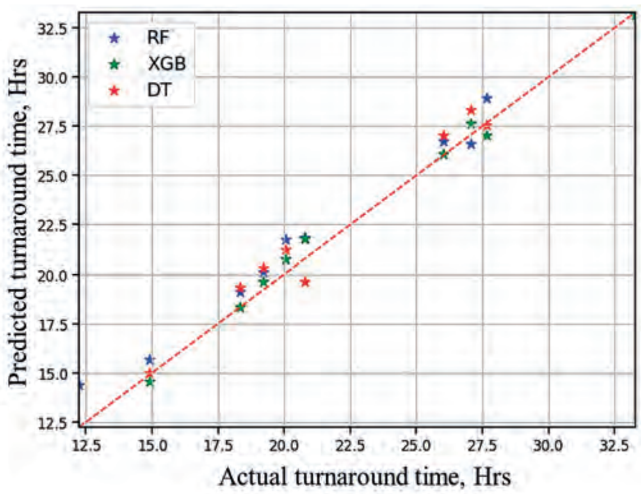


(b)

Fig. 4. Results from the training phase for the turnaround time model: (a) observed vs predicted values; (b) Taylor diagram



(a)



(b)

Fig. 5. Results from the testing phase for the turnaround time model: (a) observed vs predicted values; (b) Taylor diagram

Tab 3. Statistical evaluation of models

	Model training			Model testing		
	RF	XGBoost	DT	RF	XGBoost	DT
RMSE	2.25	1.29	1.44	1.44	0.5019	1.066
MAE	1.59	0.802	0.897	1.27	0.391	0.9
R^2	0.9355	0.9788	0.9735	0.944	0.9933	0.9696

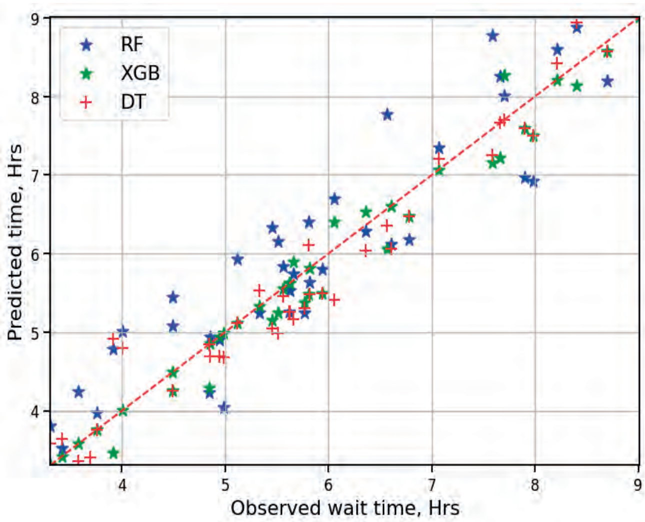
Table 3 shows the performance metrics for the three ML techniques, RF, XGBoost, and DT, in both the training and testing phases. At the model training stage, XGBoost has the lowest RMSE of 1.29, signifying its ability to minimise the average difference between the predicted and actual values. DT has a slightly higher RMSE of 1.44, whereas RF has a higher RMSE of 2.25. Furthermore, XGBoost has the lowest MAE at 0.802, suggesting smaller absolute errors on average when compared to RF (MAE = 1.59) and DT (MAE = 0.897). Moreover, XGBoost achieves the highest R^2 value of 0.9788, suggesting a superior fit to the data during training in comparison to RF (R^2 = 0.9355) and DT (R^2 = 0.9735).

When evaluated on the test dataset, XGBoost continues to demonstrate exceptional performance by achieving an RMSE of 0.5019, signifying minimal deviation between the predicted and actual values during testing. DT has a higher RMSE of 1.066, and RF has an RMSE of 1.44. In the same vein, XGBoost has the lowest MAE at 0.391, showing the smallest absolute errors on average during testing, while DT is the second-best model with a value of 0.9, and RF has a value of 1.27. Moreover, XGBoost gives the highest R^2 value of 0.9933 at the testing stage, suggesting the most accurate fit to the test data, with RF and DT achieving R^2 values of 0.944 and 0.9696, respectively.

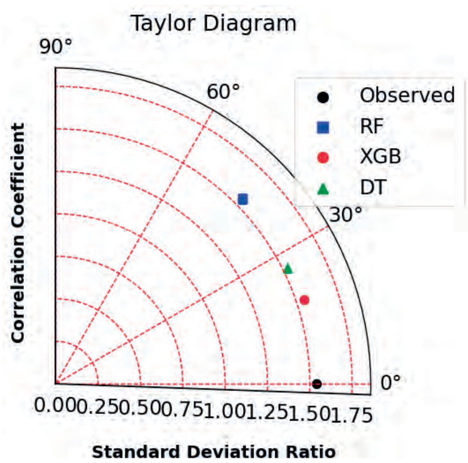
Overall, XGBoost consistently surpasses the RF and DT models in terms of RMSE, MAE, and R^2 values across both the training and testing phases. XGBoost shows exceptional predictive accuracy and model fit to the data, indicating that this is the best choice for this predictive modeling task. It is important to take into account the particular needs and limitations of the problem domain when choosing the best ML approach for real-world scenarios [100,101]

Waiting Period Model

After creating the prediction models, they were tested using data for accuracy. The model output during the training and testing phases is shown in **Figs. 6a** and **7a**. From a comparison of the models, we see that XGBoost outperforms DT and RF. Comparisons between the models were made using Taylor diagrams, as shown in **Figs. 6b** and **7b,** as this made it simple to identify the best-performing ones. Some statistical values are provided in **Table 4**.

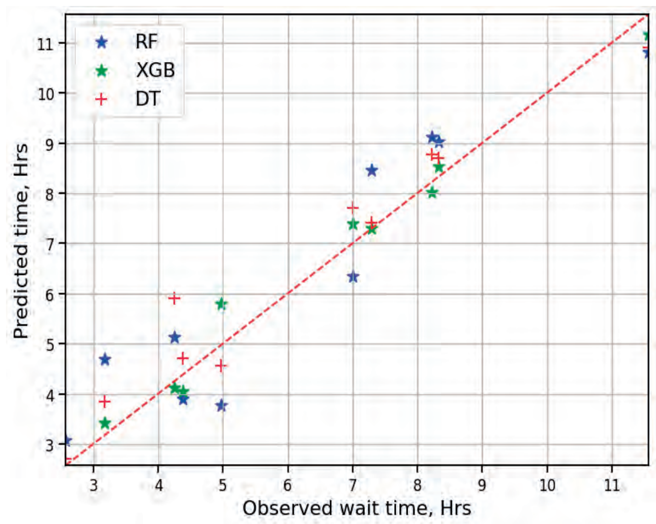


(a)

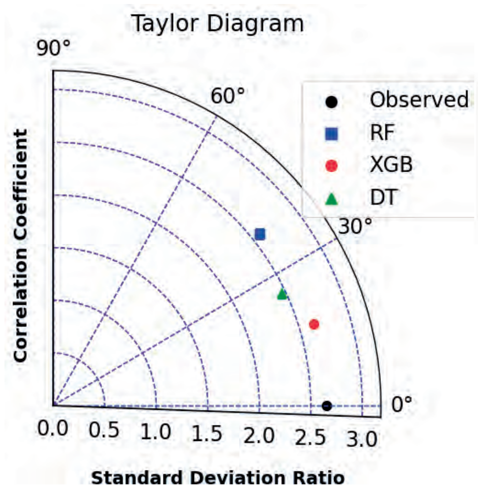


(b)

Fig. 6. Results from the training phase for the waiting period model; (a) observed vs predicted values; (b) Taylor diagram



(a)



(b)

Fig. 7. Results from the training phase for the waiting period model: (a) observed vs predicted values; (b) Taylor diagram

Tab. 4. Statistical evaluation of models

	Model training			Model testing		
	RF	XGBoost	DT	RF	XGBoost	DT
RMSE	0.618129	0.28122	0.373758	0.92869	0.394403	0.697345
MAE	0.51798	0.193347	0.304867	0.873883	0.327297	0.558989
R^2	0.838893	0.966654	0.941097	0.877256	0.977862	0.930792

It can be observed from **Table 4** that in regard to RMSE values, XGBoost (0.28122) performs better than both RF (0.618129) and DT (0.373758) by generating more precise predictions with lower error. DT has a slightly higher RMSE compared to XGBoost, and RF has the highest RMSE of the three techniques. The results for the MAE values are also consistent, with XGBoost having the lowest error at 0.193347, followed by DT at 0.304867, and RF at 0.51798. Once more, XGBoost yields exceptional performance in terms of reducing prediction errors, this result is also agreed by other researchers [102,103].

From the R^2 values, we see that XGBoost achieves an impressive score of 0.966654, indicating that it can explain a larger proportion of the variance in the data compared to the other two techniques. RF also achieves good results, with an R^2 of 0.838893, suggesting a strong fit to the data. Although DT is still effective, it has a slightly lower R^2 of 0.941097 in comparison to XGBoost and RF. Overall, XGBoost stands out as the best technique across all three metrics, demonstrating its effectiveness in predictive modeling tasks. It attains the lowest errors (RMSE and MAE) and the highest R^2 value, showcasing superior predictive performance and a better fit to the data. RF also demonstrates strong performance, especially in R^2, but falls slightly short of XGBoost in error metrics. When it comes to delivering results, DT lags behind XGBoost and RF, especially in terms of prediction accuracy. Hence, XGBoost seems to be the most appropriate option out of the three techniques for the specified predictive modeling task.

Challenges and Obstacles

When ML is applied to optimise port operations, which are essential for boosting efficiency, lowering costs, and improving overall performance, both obstacles and possibilities arise, for example:

- **Data quality and availability:** One of the most significant challenges that needs to be overcome in order to successfully use ML-based models for the prediction of port operations involves the availability and quality of test data. Despite the fact that data from an AIS and other sources provide essential information, these sources may be inconsistent, noisy, or incomplete. The process of cleaning and preparing such data in order to guarantee their quality and dependability can be both time-consuming and resource-intensive [104,105].
- **Dynamic complexity:** Port operations entail complex operations involving a variety of elements, such as the arrival and departure of vessels, timetables for berthing, handling of cargo, and the conditions of the environment [106]. To effectively model these intricate dynamics, powerful ML methods that are able to capture nonlinear interactions and temporal dependencies are required [107,108].
- **Scalability:** This is a concern for ML models since they process massive amounts of data in real time, which gives rise to scalability issues. In order to implement these systems in practice, it is vital to deploy ML algorithms that are capable of effectively managing such large amounts of data while maintaining real-time responsiveness [109,110]such as aircraft design, wind turbines, and heat transfer. Each airfoil has different aerodynamic coefficients. Obtaining the aerodynamic coefficients is a must to optimize the airfoil design. Engineers use various methods to get the airfoil aerodynamic coefficients. A prediction method is an approximation approach that effectively reduces time and cost. This article uses convolutional neural networks (CNN.
- **Interpretability:** ML models, and deep learning algorithms in particular, frequently meet with criticism due to their inability to be interpreted. In port operations, where decisions can have enormous repercussions, stakeholders demand models that are both visible and interpretable, in order to comprehend the reasoning behind projections and to make decisions that are based on accurate information [111,112]especially online news. They just get news and are unable to filter out inappropriate stuff. The media website conveys a great deal of information. Popular news websites are one source for keeping up with the newest news. It requires a significant amount of work to deliver news on prominent websites and to choose content that is not incorrect. To crawl the web and analyse enormous data, massive computer power is required, and solutions to lower the process's space and temporal complexity must be created.Data mining is seen to be a solution to the aforementioned difficulties since it extracts particular information based on defined attributes. This research investigated a model to determine the content of false news information in Indonesian popular news. Firstly, preprocessing process from dataset that collected from keaggle. Secondly, we try use classification methods to determined which the optimal method to classify fake news. Thirdly, we use another public dataset for testing method. Furthermore, five machine learning classifiers are compared: Support Vector Machine (SVM.

Prospects

- **Predictive analytics:** ML makes it possible to apply predictive analytics to a variety of port operations, such as the arrival times of vessels, the distribution of berths, the processing of cargo, and turnaround times. In order to anticipate future occurrences and optimise resource allocation, ML models may analyse past data as well as information that is collected in real time. This results in increased efficiency and decreased delays [9,113]Carbon Monoxide (CO.
- **Optimisation:** The optimisation of vessel scheduling, berth usage, and resource allocation may be improved by the application of ML-based optimisation approaches, which can also improve port operations. The ability of ML algorithms to develop optimum solutions that strike a balance between conflicting demands arises because they take into account many restrictions and objectives, such as lowering waiting times, optimising throughput, and reducing environmental effects [114,115].
- **Automation:** ML makes it possible to automate regular jobs as part of port operations, such as inspecting cargo, tracking containers, and scheduling repairs. Autonomous systems that are equipped with ML algorithms have the ability to increase safety by recognising possible dangers and managing risks, as well as expediting operations and minimising the amount of manual work required.

- **Decision support:** Decision support systems based on ML can offer port operators, terminal managers, and other stakeholders useful insights and suggestions. These systems provide assistance to decision makers in the process of making informed choices, improving operations, and limiting risks. They do this by evaluating data from a variety of sources and simulating numerous situations.

The conclusion may be drawn that although ML-based model prediction may pose problems in terms of data quality, complexity, scalability, and interpretability, it also presents a valuable opportunity for improving port operations through predictive analytics, optimisation, automation, and decision support. Ports can capture the full potential of ML by tackling these difficulties and capitalising on the possibilities. This will allow them to enhance efficiency, cut costs, and maintain their competitive edge in the global marine business.

CONCLUSION

The aim of this study was to create a port management system based on powerful prediction and classification algorithms that were capable of accurately projecting ship stay durations and delays. Our study contributes to maritime logistics research by addressing an important gap in existing port analytic frameworks. The suggested method not only helps with port decision-making but also predicts service interruptions, thereby improving overall port efficiency. To demonstrate the efficacy of the proposed technique, we conducted a case study with data from a modern port, and used feature analysis to identify the primary aspects influencing maritime logistics, resulting in a better knowledge of port operational complexity. The XGBoost regressor, trained using Bayesian-optimised hyperparameters, emerged as the best-performing model. Its strong gradient-boosting architecture effectively extracted subtle patterns from the data, yielding extremely accurate prediction models. We also trained DT and RF regressors for comparison with the XGBoost model; whereas DT is a simple solution, RF addresses overfitting and improves forecast accuracy by combining several distinct trees.

On both the training and testing datasets, the XGBoost model consistently outperformed the other approaches in terms of RMSE, MAE, and R^2 values for both the turnaround time and waiting period models. For the turnaround time, the XGBoost model had the lowest RMSE of 1.29 during training and 0.5019 during testing, as well as the lowest MAE values of 0.802 during training and 0.391 during testing. It also had the greatest R^2 values of 0.9788 during training and 0.9933 during testing. Similarly, when used in the waiting period model, XGBoost outperformed RF and DT with the lowest RMSE, MAE, and greatest R^2 values in both the training and testing phases.

Overall, our findings highlight the importance of AI-driven techniques in terms of transforming port administration and promoting innovation in marine logistics. Ports can improve efficiency, reduce costs, and remain competitive in the global marine business by utilising modern predictive modeling approaches. When deciding on the best ML method, it is critical to examine the unique demands and restrictions of each port business.

ACKNOWLEDGEMENT

Financial support for this research was provided by Ho Chi Minh University of Transport, Ho Chi Minh city, Vietnam.

REFERENCES

1. Pham NDK, Dinh GH, Pham HT, Kozak J, Nguyen HP. Role of Green Logistics in the Construction of Sustainable Supply Chains. Polish Marit Res 2023;30:191–211. https://doi.org/10.2478/pomr-2023-0052.

2. Nguyen HP, Nguyen CTU, Tran TM, Dang QH, Pham NDK. Artificial Intelligence and Machine Learning for Green Shipping: Navigating towards Sustainable Maritime Practices. JOIV Int J Informatics Vis 2024;8:1–17. https://doi.org/10.62527/joiv.8.1.2581.

3. Yalama V, Yakovleva O, Trandafilov V, Khmelniuk M. Future Sustainable Maritime Sector: Energy Efficiency Improvement and Environmental Impact Reduction for Fishing Carriers Older than 20 Years in the Fleet Part II. Polish Marit Res 2022;29:78–88. https://doi.org/10.2478/pomr-2022-0028.

4. Vakili S, Ölçer AI, Schönborn A, Ballini F, Hoang AT. Energy-related clean and green framework for shipbuilding community towards zero-emissions: A strategic analysis from concept to case study. Int J Energy Res 2022;46:20624–49. https://doi.org/10.1002/er.7649.

5. Gupta P, Rasheed A, Steen S. Ship performance monitoring using machine-learning. Ocean Eng 2022;254:111094. https://doi.org/10.1016/j.oceaneng.2022.111094.

6. Lee H, Chatterjee I, Cho G. AI-Powered Intelligent Seaport Mobility: Enhancing Container Drayage Efficiency through Computer Vision and Deep Learning. Appl Sci 2023;13:12214.

7. Farzadmehr M, Carlan V, Vanelslander T. Contemporary challenges and AI solutions in port operations: applying Gale–Shapley algorithm to find best matches. J Shipp Trade 2023;8:27.

8. Nguyen HP, Nguyen PQP, Nguyen TP. Green Port Strategies in Developed Coastal Countries as Useful Lessons for the Path of Sustainable Development: A case study in Vietnam. Int J Renew Energy Dev 2022;11:950–62. https://doi.org/10.14710/ijred.2022.46539.

9. Nguyen HP, Nguyen PQP, Nguyen DKP, Bui VD, Nguyen DT. Application of IoT Technologies in Seaport Management. JOIV Int J Informatics Vis 2023;7:228. https://doi.org/10.30630/joiv.7.1.1697.

10. Le TT, Nguyen HP, Rudzki K, Rowiński L, Bui VD, Truong TH, et al. Management Strategy for Seaports Aspiring to Green Logistical Goals of IMO: Technology and Policy Solutions. Polish Marit Res 2023;30:165–87. https://doi.org/10.2478/pomr-2023-0031.

11. Vu VV, Le PT, Do TMT, Nguyen TTH, Tran NBM, Paramasivam P, et al. An insight into the Application of AI in maritime and Logistics toward Sustainable Transportation. JOIV Int J Informatics Vis 2024;8:158–74. https://doi.org/10.62527/joiv.8.1.2641.

12. Priya JC, Rudzki K, Nguyen XH, Nguyen HP, Chotechuang N, Pham NDK. Blockchain-Enabled Transfer Learning for Vulnerability Detection and Mitigation in Maritime Logistics. Polish Marit Res 2024;31:135–45. https://doi.org/10.2478/pomr-2024-0014.

13. Nadi A, Sharma S, Snelder M, Bakri T, van Lint H, Tavasszy L. Short-term prediction of outbound truck traffic from the exchange of information in logistics hubs: A case study for the port of Rotterdam. Transp Res Part C Emerg Technol 2021;127:103111.

14. Hirata E, Watanabe D, Lambrou M, Banyai T, Banyai A, Kaczmar I. Shipping digitalization and automation for the smart port. Supply Chain Adv New Perspect Ind 40 Era 2022.

15. Lim Y, Choi G, Lee K. A Development of Embedded Anomaly Behavior Packet Detection System for IoT Environment using Machine Learning Techniques. Int J Adv Sci Eng Inf Technol 2020;10:1340–5. https://doi.org/10.18517/ijaseit.10.4.12762.

16. Kimera D, Nangolo FN. Improving ship yard ballast pumps' operations: A PCA approach to predictive maintenance. Marit Transp Res 2020;1:100003. https://doi.org/10.1016/j.martra.2020.100003.

17. Flaieh EH, Hamdoon FO, Jaber AA. Estimation the Natural Frequencies of a Cracked Shaft Based on Finite Element Modeling and Artificial Neural Network. Int J Adv Sci Eng Inf Technol 2020;10:1410–6. https://doi.org/10.18517/ijaseit.10.4.12211.

18. Wrzask K, Kowalski J, Le VV, Nguyen VG, Cao DN. Fault detection in the marine engine using a support vector data description method. J Mar Eng Technol 2024:1–11. https://doi.org/10.1080/20464177.2024.2318844.

19. Zaman A, Ren B, Liu X. Artificial Intelligence-Aided Automated Detection of Railroad Trespassing. Transp Res Rec J Transp Res Board 2019;2673:25–37. https://doi.org/10.1177/0361198119846468.

20. Tsolakis N, Zissis D, Papaefthimiou S, Korfiatis N. Towards AI driven environmental sustainability: an application of automated logistics in container port terminals. Int J Prod Res 2022;60:4508–28. https://doi.org/10.1080/00207543.2021.1914355.

21. Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manage 2021;57:101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

22. Enholm IM, Papagiannidis E, Mikalef P, Krogstie J. Artificial Intelligence and Business Value: a Literature Review. Inf Syst Front 2022;24:1709–34. https://doi.org/10.1007/s10796-021-10186-w.

23. Jiang H, Xiong W, Cao Y. A Conceptual Model of Excellent Performance Mode of Port Enterprise Logistics Management. Polish Marit Res 2017;24:34–40. https://doi.org/10.1515/pomr-2017-0102.

24. Iris Ç, Lam JSL. A review of energy efficiency in ports: Operational strategies, technologies and energy management systems. Renew Sustain Energy Rev 2019;112:170–82.

25. Attanasio G, Battistella C, Chizzolini E. The future of energy management: Results of a Delphi panel applied in the case of ports. J Clean Prod 2023;417:137947.

26. Yau K-LA, Peng S, Qadir J, Low Y-C, Ling MH. Towards Smart Port Infrastructures: Enhancing Port Activities Using Information and Communications Technology. IEEE Access 2020;8:83387–404. https://doi.org/10.1109/ACCESS.2020.2990961.

27. Molavi A, Lim GJ, Race B. A framework for building a smart port and smart port index. Int J Sustain Transp 2020;14:686–700. https://doi.org/10.1080/15568318.2019.1610919.

28. Hoang AT, Foley AM, Nižetić S, Huang Z, Ong HC, Ölçer AI, et al. Energy-related approach for reduction of CO2 emissions: A strategic review on the port-to-ship pathway. J Clean Prod 2022;355:131772. https://doi.org/10.1016/j.jclepro.2022.131772.

29. Wang B, Liu Q, Wang L, Chen Y, Wang J. A review of the port carbon emission sources and related emission reduction technical measures. Environ Pollut 2023;320:121000.

2022;21:73–81. https://doi.org/10.1080/20464177.2019.1595355.

30. Sinha D, Roy Chowdhury S. A framework for ensuring zero defects and sustainable operations in major Indian ports. Int J Qual Reliab Manag 2022;39:1896–936.

31. Chu Z, Yan R, Wang S. Vessel turnaround time prediction: A machine learning approach. Ocean Coast Manag 2024;249:107021.

32. Gucma S. Conditions of Safe Ship Operation in Seaports – Optimization of Port Waterway Parameters. Polish Marit Res 2019;26:22–9. https://doi.org/10.2478/pomr-2019-0042.

33. Alamoush AS, Ballini F, Ölçer AI. Ports' technical and operational measures to reduce greenhouse gas emission and improve energy efficiency: A review. Mar Pollut Bull 2020;160:111508. https://doi.org/10.1016/j.marpolbul.2020.111508.

34. Alamoush AS, Ölçer AI, Ballini F. Port greenhouse gas emission reduction: Port and public authorities' implementation schemes. Res Transp Bus Manag 2022;43:100708.

35. Rudzki K, Gomulka P, Hoang AT. Optimization Model to Manage Ship Fuel Consumption and Navigation Time. Polish Marit Res 2022;29:141–53. https://doi.org/10.2478/pomr-2022-0034.

36. Hu Z, Zhou T, Zhen R, Jin Y, Li X, Osman MT. A two-step strategy for fuel consumption prediction and optimization of ocean-going ships. Ocean Eng 2022;249:110904.

37. Gao C-F, Hu Z-H. Speed optimization for container ship fleet deployment considering fuel consumption. Sustainability 2021;13:5242.

38. Lamas MI, C.G. R, J. T, J.D. R. Numerical Analysis of Emissions from Marine Engines Using Alternative Fuels. Polish Marit Res 2015;22:48–52. https://doi.org/10.1515/pomr-2015-0070.

39. Hoang AT, Pandey A, Martinez De Osés FJ, Chen W-H, Said Z, Ng KH, et al. Technological solutions for boosting hydrogen role in decarbonization strategies and net-zero goals of world shipping: Challenges and perspectives. Renew Sustain Energy Rev 2023;188:113790. https://doi.org/10.1016/j.rser.2023.113790.

40. Zeńczak W, Gromadzińska AK. Preliminary Analysis of the Use of Solid Biofuels in a Ship's Power System. Polish Marit Res 2020;27:67–79. https://doi.org/10.2478/pomr-2020-0067.

41. Hoang AT, Tran VD, Dong VH, Le AT. An experimental analysis on physical properties and spray characteristics of an ultrasound-assisted emulsion of ultra-low-sulphur diesel and Jatropha-based biodiesel. J Mar Eng Technol

42. Kim J, Rahimi M, Newell J. Life-Cycle Emissions from Port Electrification: A Case Study of Cargo Handling Tractors at the Port of Los Angeles. Int J Sustain Transp 2012;6:321–37. https://doi.org/10.1080/15568318.2011.606353.

43. Jonathan YCE, Kader SBA. Prospect of Emission Reduction Standard for Sustainable Port Equipment Electrification. Int J Eng 2018;31. https://doi.org/10.5829/ije.2018.31.08b.25.

44. Nguyen HP, Hoang AT, Nizetic S, Nguyen XP, Le AT, Luong CN, et al. The electric propulsion system as a green solution for management strategy of CO2 emission in ocean shipping: A comprehensive review. Int Trans Electr Energy Syst 2021;31:e12580. https://doi.org/10.1002/2050-7038.12580.

45. Saether EA, Eide AE, Bjørgum Ø. Sustainability among Norwegian maritime firms: Green strategy and innovation as mediators of long-term orientation and emission reduction. Bus Strateg Environ 2021;30:2382–95.

46. Agarwala P, Chhabra S, Agarwala N. Using digitalisation to achieve decarbonisation in the shipping industry. J Int Marit Safety, Environ Aff Shipp 2021;5:161–74.

47. Serra P, Fancello G. Towards the IMO's GHG Goals: A Critical Overview of the Perspectives and Challenges of the Main Options for Decarbonizing International Shipping. Sustainability 2020;12:3220. https://doi.org/10.3390/su12083220.

48. Gupta S, Modgil S, Choi T-M, Kumar A, Antony J. Influences of artificial intelligence and blockchain technology on financial resilience of supply chains. Int J Prod Econ 2023;261:108868. https://doi.org/10.1016/j.ijpe.2023.108868.

49. Nguyen HP, Le PQH, Pham VV, Nguyen XP, Balasubramaniam D, Hoang A-T. Application of the Internet of Things in 3E (efficiency, economy, and environment) factor-based energy management as smart and sustainable strategy. Energy Sources, Part A Recover Util Environ Eff 2021:1–23. https://doi.org/10.1080/15567036.2021.1954110.

50. Xu H, Liu J, Xu X, Chen J, Yue X. The impact of AI technology adoption on operational decision-making in competitive heterogeneous ports☆. Transp Res Part E Logist Transp Rev 2024;183:103428. https://doi.org/10.1016/j.tre.2024.103428.

51. Koh L, Dolgui A, Sarkis J. Blockchain in transport and logistics – paradigms and transitions. Int J Prod Res

2020;58:2054–62. https://doi.org/10.1080/00207543.2020.1736428.

52. Lambert N, Turner J, Hamflett A. Technology and the blue economy: from autonomous shipping to big data. Kogan Page Publishers; 2019.

53. Tu H, Xia K, Zhao E, Mu L, Sun J. Optimum trim prediction for container ships based on machine learning. Ocean Eng 2023;277:111322. https://doi.org/10.1016/j.oceaneng.2022.111322.

54. Senol YE, Seyhan A. A novel machine-learning based prediction model for ship manoeuvring emissions by using bridge simulator. Ocean Eng 2024;291:116411. https://doi.org/10.1016/j.oceaneng.2023.116411.

55. Bassam AM, Phillips AB, Turnock SR, Wilson PA. Ship speed prediction based on machine learning for efficient shipping operation. Ocean Eng 2022;245:110449. https://doi.org/10.1016/j.oceaneng.2021.110449.

56. Walker AM, Cliff A, Romero J, Shah MB, Jones P, Felipe Machado Gazolla JG, et al. Evaluating the performance of random forest and iterative random forest based methods when applied to gene expression data. Comput Struct Biotechnol J 2022;20:3372–86. https://doi.org/10.1016/j.csbj.2022.06.037.

57. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst Appl 2019;134:93–101. https://doi.org/10.1016/j.eswa.2019.05.028.

58. Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J Promot Commun Stat Stata 2020;20:3–29. https://doi.org/10.1177/1536867X20909688.

59. Breiman L. Random Forests. Mach Learn 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.

60. Talebi H, Peeters LJM, Otto A, Tolosana-Delgado R. A Truly Spatial Random Forests Algorithm for Geoscience Data Analysis and Modelling. Math Geosci 2022;54:1–22. https://doi.org/10.1007/s11004-021-09946-w.

61. Gholizadeh M, Jamei M, Ahmadianfar I, Pourrajab R. Prediction of nanofluids viscosity using random forest (RF) approach. Chemom Intell Lab Syst 2020;201:104010. https://doi.org/10.1016/J.CHEMOLAB.2020.104010.

62. Chen P, Niu A, Jiang W, Liu D. Air Pollutant Prediction: Comparisons between LSTM, Light GBM and Random Forest. Geophys Res Abstr 2019;21.

63. Nguyen G V., Le X-H, Van LN, Jung S, Yeon M, Lee G. Application of Random Forest Algorithm for Merging Multiple Satellite Precipitation Products across South Korea. Remote Sens 2021;13:4033. https://doi.org/10.3390/rs13204033.

64. Said Z, Sharma P, Bora BJ, Pandey AK. Sonication impact on thermal conductivity of f-MWCNT nanofluids using XGBoost and Gaussian process regression. J Taiwan Inst Chem Eng 2023;145:104818. https://doi.org/10.1016/j.jtice.2023.104818.

65. Kumar K P, Alruqi M, Hanafi HA, Sharma P, Wanatasanappan VV. Effect of particle size on second law of thermodynamics analysis of Al2O3 nanofluid: Application of XGBoost and gradient boosting regression for prognostic analysis. Int J Therm Sci 2024;197:108825. https://doi.org/10.1016/j.ijthermalsci.2023.108825.

66. Zou M, Jiang W-G, Qin Q-H, Liu Y-C, Li M-L. Optimized XGBoost Model with Small Dataset for Predicting Relative Density of Ti-6Al-4V Parts Manufactured by Selective Laser Melting. Materials (Basel) 2022;15:5298. https://doi.org/10.3390/ma15155298.

67. Qiu Y, Zhou J, Khandelwal M, Yang H, Yang P, Li C. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Eng Comput 2022;38:4145–62. https://doi.org/10.1007/s00366-021-01393-9.

68. Chen T, Guestrin C. XGBoost. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., New York, NY, USA: ACM; 2016, p. 785–94. https://doi.org/10.1145/2939672.2939785.

69. Said Z, Sohail M, Tiwari AK. Recent advances in machine learning research for nanofluid heat transfer in renewable energy. Adv. Nanofluid Heat Transf., Elsevier; 2022, p. 203–28. https://doi.org/10.1016/B978-0-323-88656-7.00011-8.

70. Nguyen VN, Tarełko W, Sharma P, El-Shafay AS, Chen W-H, Nguyen PQP, et al. Potential of Explainable Artificial Intelligence in Advancing Renewable Energy: Challenges and Prospects. Energy & Fuels 2024;38:1692–712. https://doi.org/10.1021/acs.energyfuels.3c04343.

71. Jamei M, Sharma P, Ali M, Bora BJ, Malik A, Paramasivam P, et al. Application of an explainable glass-box machine learning approach for prognostic analysis of a biogas-powered small agriculture engine. Energy 2024;288:129862. https://doi.org/10.1016/j.energy.2023.129862.

72. Hafeez MA, Rashid M, Tariq H, Abideen ZU, Alotaibi SS, Sinky MH. Performance Improvement of Decision Tree: A Robust Classifier Using Tabu Search Algorithm. Appl Sci 2021;11:6728. https://doi.org/10.3390/app11156728.

73. Nanfack G, Temple P, Frénay B. Constraint Enforcement on Decision Trees: A Survey. ACM Comput Surv 2022;54:1–36. https://doi.org/10.1145/3506734.

74. Kotsiantis SB. Decision trees: a recent overview. Artif Intell Rev 2013;39:261–83. https://doi.org/10.1007/s10462-011-9272-4.

75. Custode LL, Iacca G. Evolutionary Learning of Interpretable Decision Trees. IEEE Access 2023;11:6169–84. https://doi.org/10.1109/ACCESS.2023.3236260.

76. Said Z, Sharma P, Tiwari AK, Le VV, Huang Z, Bui VG, et al. Application of novel framework based on ensemble boosted regression trees and Gaussian process regression in modelling thermal performance of small-scale Organic Rankine Cycle (ORC) using hybrid nanofluid. J Clean Prod 2022;360:132194. https://doi.org/10.1016/j.jclepro.2022.132194.

77. Sharma P, Sahoo BB, Said Z, Hadiyanto H, Nguyen XP, Nižetić S, et al. Application of machine learning and Box-Behnken design in optimizing engine characteristics operated with a dual-fuel mode of algal biodiesel and waste-derived biogas. Int J Hydrogen Energy 2023;48:6738–60. https://doi.org/10.1016/j.ijhydene.2022.04.152.

78. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 2021;7:e623. https://doi.org/10.7717/peerj-cs.623.

79. Sharma P, Sharma AK. AI-Based Prognostic Modeling and Performance Optimization of CI Engine Using Biodiesel-Diesel Blends. Int J Renew Energy Resour 2021;11:701–8.

80. Pozzi F, Di Matteo T, Aste T. Exponential smoothing weighted correlations. Eur Phys J B 2012;85:175. https://doi.org/10.1140/epjb/e2012-20697-x.

81. Dodo UA, Ashigwuike EC, Abba SI. Machine learning models for biomass energy content prediction: A correlation-based optimal feature selection approach. Bioresour Technol Reports 2022;19:101167. https://doi.org/10.1016/j.biteb.2022.101167.

82. Le AT, Pandey A, Sirohi R, Sharma P, Chen W-H, Pham NDK, et al. Precise Prediction of Biochar Yield and Proximate Analysis by Modern Machine Learning and SHapley Additive exPlanations. Energy & Fuels 2023;37:17310–17327. https://doi.org/10.1021/acs.energyfuels.3c02868.

83. Rosiani D, Gibral Walay M, Rahalintar P, Candra AD, Sofyan A, Arison Haratua Y. Application of Artificial Intelligence in Predicting Oil Production Based on Water Injection Rate. Int J Adv Sci Eng Inf Technol 2023;13:2338–44. https://doi.org/10.18517/ijaseit.13.6.19399.

84. Kim S-W. Change in Attitude toward Artificial Intelligence through Experiential Learning in Artificial Intelligence Education. Int J Adv Sci Eng Inf Technol 2023;13:1953–9. https://doi.org/10.18517/ijaseit.13.5.19039.

85. Kim S-W, Go H, Hong S-J, Lee Y. An Approach to the Utilization of Design Thinking in Artificial Intelligence Education. Int J Adv Sci Eng Inf Technol 2023;13:2198–204. https://doi.org/10.18517/ijaseit.13.6.19042.

86. Feng Y, Wu Q. A statistical learning assessment of Huber regression. J Approx Theory 2022;273:105660. https://doi.org/10.1016/j.jat.2021.105660.

87. Meyers SD, Azevedo L, Luther ME. A Scopus-based bibliometric study of maritime research involving the Automatic Identification System. Transp Res Interdiscip Perspect 2021;10:100387. https://doi.org/10.1016/j.trip.2021.100387.

88. Lee E, Mokashi AJ, Moon SY, Kim G. The Maturity of Automatic Identification Systems (AIS) and Its Implications for Innovation. J Mar Sci Eng 2019;7:287. https://doi.org/10.3390/jmse7090287.

89. Goudosis A, Katsikas S. Secure Automatic Identification System (SecAIS): Proof-of-Concept Implementation. J Mar Sci Eng 2022;10:805. https://doi.org/10.3390/jmse10060805.

90. Wang Z, Xia L, Yuan H, Srinivasan RS, Song X. Principles, research status, and prospects of feature engineering for data-driven building energy prediction: A comprehensive review. J Build Eng 2022;58:105028. https://doi.org/10.1016/j.jobe.2022.105028.

91. Berberich J, Kohler J, Muller MA, Allgower F. Data-Driven Model Predictive Control With Stability and Robustness Guarantees. IEEE Trans Automat Contr 2021;66:1702–17. https://doi.org/10.1109/TAC.2020.3000182.

92. Handari BD, Wulandari D, Aquita NA, Leandra S, Sarwinda D, Hertono GF. Comparing Restricted Boltzmann Machine â€“ Backpropagation Neural Networks, Artificial Neural Network â€“ Genetic Algorithm and Artificial Neural Network â€“ Particle Swarm Optimization for Predicting DHF Cases in DKI Jakarta. Int J Adv Sci Eng Inf Technol 2022;12:2476–84. https://doi.org/10.18517/ijaseit.12.6.16226.

93. Ayulani ID, Yunawan AM, Prihutaminingsih T, Sarwinda D, Ardaneswari G, Handari BD. Tree-Based Ensemble Methods and Their Applications for Predicting Studentsâ€™ Academic Performance. Int J Adv Sci Eng

Inf Technol 2023;13:919–27. https://doi.org/10.18517/ijaseit.13.3.16880.

94. Masmoudi K, Masmoudi A. A new class of continuous Bayesian networks. Int J Approx Reason 2019;109:125–38. https://doi.org/10.1016/j.ijar.2019.03.010.

95. Chen H, Li X, Feng Z, Wang L, Qin Y, Skibniewski MJ, et al. Shield attitude prediction based on Bayesian-LGBM machine learning. Inf Sci (Ny) 2023;632:105–29. https://doi.org/10.1016/j.ins.2023.03.004.

96. Jung Y. Multiple predicting K -fold cross-validation for model selection. J Nonparametr Stat 2018;30:197–215. https://doi.org/10.1080/10485252.2017.1404598.

97. Fushiki T. Estimation of prediction error by using K-fold cross-validation. Stat Comput 2011;21:137–46. https://doi.org/10.1007/s11222-009-9153-8.

98. Shi R, Xu X, Li J, Li Y. Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. Appl Soft Comput 2021;109:107538.

99. Zhou J, Qiu Y, Zhu S, Armaghani DJ, Khandelwal M, Mohamad ET. Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization. Undergr Sp 2021;6:506–15.

100. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Environ 2023;31:3360–79.

101. Pan S, Zheng Z, Guo Z, Luo H. An optimized XGBoost method for predicting reservoir porosity using petrophysical logs. J Pet Sci Eng 2022;208:109520.

102. Wang C-C, Kuo P-H, Chen G-Y. Machine learning prediction of turning precision using optimized xgboost model. Appl Sci 2022;12:7739.

103. Li Y, Zeng H, Zhang M, Wu B, Qin X. Global de-trending significantly improves the accuracy of XGBoost-based county-level maize and soybean yield prediction in the Midwestern United States. GIScience Remote Sens 2024;61:2349341.

104. Zhang L, Meng Q, Xiao Z, Fu X. A novel ship trajectory reconstruction approach using AIS data. Ocean Eng 2018;159:165–74. https://doi.org/10.1016/j.oceaneng.2018.03.085.

105. Susanti R, Zaini Z, Hidayat A, Alfitri N, Rusydi MI. Identification of Coffee Types Using an Electronic Nose with the Backpropagation Artificial Neural Network. JOIV Int J Informatics Vis 2023;7:659. https://doi.org/10.30630/joiv.7.3.1375.

106. Nguyen MD, Yeon KT, Rudzki K, Nguyen HP, Pham NDK. Strategies for developing logistics centers: Technological trends and policy implications. Polish Marit Res 2023;30:129–47. https://doi.org/10.2478/pomr-2023-0066.

107. Halim C, Eka Putra NG, Nugroho NA, Suhartono D. Chest X-ray Image Classification to Identify Lung Diseases Using Convolutional Neural Network and Convolutional Block Attention Module. JOIV Int J Informatics Vis 2023;7:651–8. https://doi.org/10.30630/joiv.7.3.1136.

108. Du Y, Chen Y, Li X, Schönborn A, Sun Z. Data fusion and machine learning for ship fuel efficiency modeling: Part II – Voyage report data, AIS data and meteorological data. Commun Transp Res 2022;2:100073. https://doi.org/10.1016/j.commtr.2022.100073.

109. Zikri AA, Defianti H, Hidayat W, Purqon A. Geometry Representation Effectiveness in Improving Airfoil Aerodynamic Coefficient Prediction with Convolutional Neural Network. JOIV Int J Informatics Vis 2023;7:644. https://doi.org/10.30630/joiv.7.3.1577.

110. Guo S, Huang X, Situ Y, Huang Q, Guan K, Huang J, et al. Interpretable Machine-Learning and Big Data Mining to Predict Gas Diffusivity in Metal-Organic Frameworks. Adv Sci 2023;10. https://doi.org/10.1002/advs.202301461.

111. Triyono L, Gernowo R, Prayitno P, Rahaman M, Yudantoro TR. Fake News Detection in Indonesian Popular News Portal Using Machine Learning For Visual Impairment. JOIV Int J Informatics Vis 2023;7:726–32. https://doi.org/10.30630/joiv.7.3.1243.

112. Pristyanto Y, Mukarabiman Z, Nugraha AF. Extreme Gradient Boosting Algorithm to Improve Machine Learning Model Performance on Multiclass Imbalanced Dataset. JOIV Int J Informatics Vis 2023;7:710–5. https://doi.org/10.30630/joiv.7.3.1102.

113. Andrizal -, Chadry R, Suryani AI. Embedded System Using Field Programmable Gate Array (FPGA) myRIO and LabVIEW Programming to Obtain Data Patern Emission of Car Engine Combustion Categories. JOIV Int J Informatics Vis 2018;2:56–62. https://doi.org/10.30630/joiv.2.2.50.

114. Nguyen VG, Rajamohan S, Rudzki K, Kozak J, Sharma P, Pham NDK, et al. Using Artificial Neural Networks for Predicting Ship Fuel Consumption. Polish Marit Res 2023;30. https://doi.org/10.2478/pomr-2023-0020.

115. Meng Q, Du Y, Wang Y. Shipping log data based container ship fuel efficiency modeling. Transp Res Part B Methodol 2016. https://doi.org/10.1016/j.trb.2015.11.007.