

Forecasting and Feature Analysis of Ship Fuel Consumption by Explainable Machine Learning Approaches

Nguyen Dang Khoa Pham¹ 

Gia Huy Dinh^{1,*} 

Canh Lam Nguyen²

Hai Quoc Dang¹ 

Hoang Thai Pham¹ 

Quyen Tat Nguyen¹

Minh Cong Tran³

¹ Ho Chi Minh City University of Transport, Ho Chi Minh city, Viet Nam

² The Business School, Business Innovation Department, RMIT University, Ho Chi Minh city, Viet Nam

³ Nha Trang University, Nha Trang, Viet Nam

* Corresponding author: huy.dinh@ut.edu.vn (Gia Huy Dinh)

ABSTRACT

Rising shipping emissions greatly affect greenhouse gas (GHG) levels, so precise fuel consumption forecasting is essential to reduce environmental effects. Precision forecasts using machine learning (ML) could offer sophisticated solutions that increase the fuel efficiency and lower emissions. Indeed, five ML techniques, linear regression (LR), decision tree (DT), random forest (RF), XGBoost, and AdaBoost, were used to develop ship fuel consumption models in this study. It was found that, with an R^2 of 1, zero mean squared error (MSE), and a negligible mean absolute percentage error (MAPE), the DT model suited the training set perfectly, while R^2 was 0.8657, the MSE was 56.80, and the MAPE was 16.37% for the DT model testing. More importantly, this study provided Taylor diagrams and violin plots that helped in the identification of the best-performing models. Generally, the employed ML approaches efficiently predicted the data; however, they are black-box methods. Hence, explainable machine learning methods like Shapley additive explanations, the DT structure, and local interpretable model-agnostic explanations (LIME) were employed to comprehend the models and perform feature analysis. LIME offered insights, demonstrating that the major variables impacting predictions were distance (≤ 450.88 nm) and time ($40.70 < \text{hr} \leq 58.05$). By stressing the most important aspects, LIME can help one to comprehend the models with ease.

Keywords: ship fuel consumption, fuel consumption forecasting, model prediction, machine learning technique, operation efficiency

INTRODUCTION

As underlined by the United Nations Conference on Commerce and Development, marine transportation takes centre stage in the worldwide transportation and commerce system. Maritime transport ensures the continuous flow of products, promoting worldwide economic development [1,2]. It

accounts for about 80% of the transportation of the worldwide cargo volume [3]. This vital maritime sector underlines the critical necessity to ensure the success of ship, port, and shipyard operations, from environmental and economic perspectives [4–6]. Economically, the affordability of marine transportation directly affects the worldwide availability and pricing of products [7]. Poor operations may result in greater

transportation expenses, which are eventually passed on to customers and companies, therefore influencing the whole economy. Environmentally, the stakes are similarly great since ships' fuel use is not only a substantial running cost but also a main cause of greenhouse gas (GHG) emissions [8,9]. Shipping operations are held responsible for air pollution, as they emit gases like carbon dioxide (CO₂), sulphur oxides (SO_x), nitrogen oxides (NO_x), and particulate matter [10–12]. These emissions seriously jeopardise ecosystems and human life over the long term by contributing to global warming and climate change. The International Maritime Organization (IMO)'s 2020 GHG research study shows that the shipping industry contributed to 2.89% of all emissions globally in 2018, or over 1.076 billion tons. If no steps are taken, ship-related CO₂ emissions are predicted to increase by 50% to 250% by 2025, indicating that there is an urgent need for efficient strategies of decarbonisation in the maritime sector [13–15]. At the same time, global marine fuel prices remain high, and fleet running expenses increasingly include fuel usage charges. Since most boats in the world run on marine diesel engines, their reliability has a substantial effect on ship safety and efficiency. Studying ways to improve ship energy efficiency is crucial given a strict carbon reduction target [16,17].

Adopted by the IMO, one of the most important turning points in these initiatives is the International Convention for the Prevention of Pollution from Ships (MARPOL). MARPOL covers a wide range of maritime pollution concerns, including trash, oil spills, the movement of toxic liquids and chemicals, and ship-generated air pollution [18–20]. An important addition to MARPOL was Annex VI, which focuses especially on preventing air pollution from ships. Strong rules are set in Annex VI to control NO_x, SO_x, and other ozone-depleting compounds from ship emissions. These rules were a significant first step in tackling problems with air quality and lessening the effect of shipping on the environment [21,22]. MARPOL Annex VI was then expanded to include rules specifically meant to lower GHG emissions; this modification included steps to reduce marine activity-related CO₂, methane (CH₄), nitrous oxide (N₂O), and ozone (O₃) emissions. By including these gases within MARPOL, the shipping sector's part in global climate change was underlined, as well as the need for thorough plans to reduce these emissions [23,24]. Due to this issue, the optimisation of fuel economy in marine operations is desperately needed, considering these financial and environmental consequences. Driven by incumbent rules enforcing tougher emission limits and international accords like the Paris Agreement, the marine sector is forced to look for creative ideas to lessen its environmental impact as the globe moves toward more sustainable practices [25]. This calls for the research and application of cutting-edge technology to improve sustainability and operational efficiency [26].

The marine sector is under pressure to develop creative ideas to improve fuel economy and lower emissions in response to strict rules and the urgent necessity of environmental sustainability [27]. The acceptance of cutting-edge technology such as artificial intelligence (AI) and machine learning

(ML) is becoming widespread. For predictive modelling, fuel economy optimisation, and general operational efficiency enhancements, these technologies provide strong tools. ML algorithms can provide accurate forecasts and useful insights by examining large volumes of data from many sources, including real-time ship performance data, meteorological conditions, and past trip information. ML models may, for example, optimise ship routes to evade bad weather, save fuel, and lower pollutants. Furthermore, predictive maintenance made possible by AI can assist in spotting possible equipment breakdowns before they start, therefore lowering downtime and increasing ship operation efficiency. Including this technology in marine activities not only helps to comply with environmental rules but also results in major cost savings and improved industrial competitiveness. Several authors have implemented ML approaches to improve and optimise ship fuel consumption for better ship management and economic management. Zhang et al. [28] employed the deep learning approach for the prediction of ship fuel consumption. The data were gathered using different approaches like voyage reports, sensors, and hydrometeorological data, corresponding to 266 variables obtained during sea trials. The fundamental linkages in the given dataset were discovered using a variable importance estimate approach employing a decision tree. Therefore, a deep learning model was created to grasp the effect of the sailing speed, headings, draft/displacement, trim, weather, ocean conditions, etc., on ship fuel consumption. This was accomplished in bi-directional long short-term memory (Bi-LSTM) networks by including an attention mechanism. Training data streams matching real ship fuel consumption, as well as internal and exterior operational variables, helped to show the promise of the new approach. Agand et al. [29] employed a comparative approach using ML and in-service data for the prediction of the fuel consumption of passenger ferries. Based on data from a passenger ship, the study proposed methods for predicting fuel consumption. The employed methods helped in the reduction of overfitting, missing data, and multicollinearity. The study examines multiple prediction models, including multiple linear regression (MLR), a decision tree (DT), an artificial neural network (ANN), and ensemble techniques. The most accurate model was created using XGBoost, a boosting ensemble strategy. Chen et al. [30] used ML for ship fuel consumption in the case of harbour vessels. The prediction models took into account the influence of several different equilibrium levels in the dataset. To demonstrate good predictive capabilities, the historical fuel data for consumption were employed. The models that are better suited for the various time stages were found. The ability of LSTM and Bi-LSTM to make accurate predictions was validated through the process of model generalisation. Beginning at step 30, the mean absolute percentage error (MAPE) of Bi-LSTM is 3.8%, while the MAPE of EEMD-LSTM is 1.9% from step 30 to step 60. Yan et al. [31] used a two-stage ship fuel consumption forecast and reduction model for dry bulk ships. The first stage proposes and validates a fuel consumption forecasting model grounded on random forest (RF). Based on the ship

sailing speed, total cargo weight, and sea and weather conditions, the forecast model generates the hourly fuel consumption of the main engine. The RF approach has an MAPE of 7.91%. Following the forecasting model suggested in the first step, a speed optimisation model is constructed in the second stage, ensuring the expected arrival time to the target port. The suggested approach can lower ship fuel usage by 2–7%, according to numerical experiments on two consecutive eight-day cruises. Reduced CO₂ emissions will also follow from decreased ship fuel usage. Rudzki et al. [32] developed a computer-aided system that helps the ship's propulsion system select the desired outputs. The two-criteria optimisation model of this system was efficient since it enables the rational control of the navigation time and ship fuel consumption.

This research aims to improve the accuracy of fuel consumption projections, thereby addressing the major influence of growing maritime emissions. Using five machine learning methods, including linear regression (LR), DT, RF, XGBoost, and AdaBoost, this paper creates models for ship fuel consumption prediction, in which 30% of the data are set aside for testing, and 70% are used for training. In addition, using measures like R², the MAPE, and the mean squared error (MSE) for both training and testing datasets, the paper seeks to construct and assess the performance of these machine learning models. More importantly, local interpretable model-agnostic explanations (LIME) are also used to look for important factors that greatly affect fuel consumption projections. The paper stresses the need for variables like distance and time in the predictive accuracy of the models by offering insights on the practical applicability of these models, demonstrating the utility of advanced machine learning techniques in producing dependable and interpretable models for lowering environmental impacts through better fuel consumption forecasting in maritime operations.

MATERIALS AND METHODS

DATA COLLECTION

Analysing marine energy efficiency and emissions depends critically on data collected on ship fuel consumption from ship registers. Ship registries preserve thorough records on vessel characteristics, the operating history, and fuel use. Extensive information about fuel types, the quantity of fuel consumed, and operational conditions, including routes, speeds, and durations, is gathered using logbooks, onboard monitoring systems, and automated reporting systems such as the Automatic Identification System (AIS); these methods can all help to compile this information. The development of models to estimate fuel consumption and evaluate the environmental effect of shipping operations depends on accurate data collection. It also helps with regulatory compliance with global norms such as the IMO's MARPOL Annex VI, which calls on ships to cut SO_x and other

pollutants. Examining this information helps to promote the general objective of sustainable marine operations employing improved route planning, more efficient propulsion systems, and greener fuels. The raw data given in Table 1 were employed in this study.

Tab. 1. Data employed for this study

Time, hrs	Distance, nm	LFO, gallons	Time, hrs	Distance, nm	LFO, gallons
45.1	336.5	26.362	55.4	588	65.401
40.7	449.1	37.511	64.5	754.4	42.248
59.3	640.3	37.026	34.5	400.3	24.598
64	650	44.161	39.5	424.5	31.196
54.3	652.5	33.424	67.7	758	43.782
70.7	782	44.953	77.3	947.2	49.999
94.5	1103	54.939	20	207.7	14.797
71.7	647.1	47.521	70.9	764.9	52.299
54.9	633.3	34.484	53	584.9	35.256
56.3	615.3	34.479	58.4	653.6	42.012
72.3	792	42.409	71	817.6	47.379
52.3	644	33.056	37.5	438.6	24.449
40.4	450.4	29.09	24	250.3	17.993
40.6	449.3	27.06	21.3	248	13.374
71.3	808.4	48.904	40.7	452.7	32.582
38.3	436	25.034	36.2	463.6	25.576
56.6	646.5	35.902	50	637.3	33.032
64.2	743	39.762	71.1	656.2	54.951
28.8	275.7	20.221	51.8	647.3	43.057
26.4	306	22.079	64.4	650.2	40.821
57.6	623	36.555	50.8	652.1	35.44
50.1	588	30.732	63.6	675.6	42.718
54.6	649	37.398	46	448.6	29.759
40.7	464.7	32.326	37.2	454.5	32.881
44.3	449.8	30.222	87.4	743.2	64.048
29.2	347	25.965	39.8	451.7	31.221
17.9	166	14.337	36.9	444.9	28.132
10.5	60	24.057	44.6	578.1	29.767
98.9	1137	78.127	50.8	569.7	39.813
85	988	70.798	59.4	644.9	39.886
23.5	199	25.174	60.6	645.5	44.786
16.2	61.3	45.392	58.3	749.4	60.002
12.2	78	14.76	65.2	742.2	48.253
12.2	72	17.83	59.8	659.5	47.4
78.1	705.1	59.925	92.5	1094.3	75.002
71.1	716	67.44	107.4	1219.9	81.354
99.6	1117.8	83.657	105	1127.8	74.006
99.1	1128.2	83.954	93.4	1160.8	65.446
5.6	42.7	14.64	95.5	1095.2	67.44
8.1	70	24.711	106.3	1163.1	65.005
93.4	1132.3	68.643	103	1155	77.062
53.8	668.9	40.753	2.6	13.5	8.102

Time, hrs	Distance, nm	LFO, gallons	Time, hrs	Distance, nm	LFO, gallons
64.1	696	52.724	80	674.9	52.788
72.4	742.5	60.188	55.9	648	35.405
45.7	469.9	34.492	60.3	655	47.228
85.3	746.4	54.671	57.8	454	42.667
62.8	749.4	43.672	38.1	448.2	27.806
78.7	745.6	55.11	83.8	749.4	55.852
65.5	747	41.198	68.6	754	47.423
57.1	625.2	43.448	41.8	450.9	28.191
65	741.1	42.555	39	450.8	32.438
52	642.8	32.883	61.5	742.6	52.406
101.7	1196.8	76.055	62.4	746.4	43.791
91.2	1095.9	55.554	57.3	652.5	40.393
91.7	1110.1	53.932	81.5	644.6	55.643
65.9	656.5	45.368	72	744	46.058

MACHINE LEARNING

Linear regression

Linear regression (LR) is a fundamental or basic statistical technique used for modelling the connection between a dependent variable and one or more independent variables. LR seeks to project the value of the dependent variable based on the input characteristics by fitting a linear equation to observed data. Minimising the sum of squared residuals, the model guarantees the best-fit line across the data points. For linear connections, LR is strong despite its simplicity; it also provides a benchmark for more complicated models. It is less successful in identifying non-linear patterns in data, as it supposes a linear correlation [33,34].

Random forest

Random forest (RF) is an ensemble model training approach that is created with multiple decision trees. Every tree is constructed with a random feature set and a bootstrap sample of the data [35,36]. By averaging the forecasts from every individual tree, the last prediction is derived, hence improving the accuracy and resilience. RF manages high-dimensional data effectively and reduces the overfitting that is common in single decision trees. It provides a great prediction performance and robustness to noisy data, and it shines at identifying intricate patterns and interactions inside data [37].

AdaBoost

Adaptive boosting (AdaBoost) is a robust ensemble technique. It forms a strong prediction model by aggregating the outputs of numerous weak learners. Weak learners may refer to decision stumps, as AdaBoost highlights challenging situations in subsequent rounds. It tweaks the weights of misclassified occurrences repeatedly. Every weak learner receives sequential training based on the past errors of other learners. By pooling the weighted predictions of every learner, the final model increases accuracy [38,39].

Extreme gradient boosting

XGBoost is a quite scalable and highly efficient method of gradient boosting. Usually consisting of decision trees, it creates an ensemble of weak learners whereby each tree optimises a differentiable loss function, hence correcting the faults of subsequent trees. XGBoost uses sophisticated tree pruning, parallel processing, and cache optimisation to improve the speed and performance and regularising methods to prevent overfitting [40,41].

Decision tree

Applied for both classification and regression problems, the decision tree is a non-parametric supervised learning tool. It generates a tree-like model of decisions by dividing the data into subsets depending on the most important attribute at every node [42]. Every internal node stands for a feature; every branch is a decision guide; and every leaf node shows a result. Seeking to maximise information gain or decrease impurity at every split, the model recursively divides the data. Although they are interpretable, decision trees can suffer from overfitting [43,44].

MODEL EVALUATION

Coefficient of determination

The coefficient of determination (R^2) measures the proportion of the variance in the dependent variable that is predictable from the independent variable (s). It ranges from 0 to 1, with 1 indicating perfect prediction. The following expression is employed for R^2 [45]:

$$R^2 = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2}. \quad (1)$$

Mean squared error

The mean squared error (MSE) estimates the average squared difference between measured and forecasted values. It indicates the quality of the predictor by penalising larger errors more than smaller ones. The formula for the MSE is [46]

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

Mean absolute percentage error

The mean absolute percentage error (MAPE) estimates the mean absolute percentage difference between actual and predicted values, providing a relative measure of the prediction accuracy. The formula for the MAPE is [47]

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (3)$$

where y_i is the measured observation, $\hat{y}_i$ is the model-forecasted value, $\bar{y}$ is the mean of the measured observations, and n is the number of observations. Model comparison using visual approaches

RESULTS AND DISCUSSION

DATA ANALYSIS

Taylor diagrams offer a clear, all-encompassing graphical depiction of how well models replicate the observed data. Three statistical measurements, the correlation coefficient, standard deviation, and root mean square difference (RMSD), are plotted in these graphs. The correlation coefficient measures the intensity and direction of the linear link between the model predictions and data. The standard deviation shows the model predictions' relative variation compared to the observed data. The RMSD computes the magnitude of the average model prediction error [48,49] and for multiple parameters, simultaneously. Using modified Taylor diagrams has led to new conclusions about the models tested in McNamara et al. (2013).

Combining elements of box plots and kernel density plots, violin plots are a flexible instrument for showing the density and distribution of data. They are especially good for contrasting the prediction distributions of several models. Every violin plot features a horizontal axis showing the density estimate and a vertical axis showing the data values [50–52]. The violin plot's form offers insights into skewness and multimodality, distribution qualities that conventional box graphs might not highlight. Furthermore, violin charts enable thorough comparisons across several models by stressing the central tendency and variability of the model forecasts. Within the framework of model comparison, violin charts help researchers investigate the concentration and distribution of forecasts, therefore enabling the identification of models with either similar or different behaviours [53,54] systematic reviews regularly indicate that single-case experimental designs (SCEDs).

The correlation heatmap for the data used is depicted in Fig. 1(a). Three variables, the time (hrs), distance (nautical miles, nm), and low-sulphur fuel oil (LFO) (gallons), have correlations shown by the correlation heat map for ship fuel consumption data. Using values ranging from -1 to 1, the heatmap shows the correlation coefficients using a colour gradient. A perfect positive correlation is indicated by a value of 1; a perfect negative correlation is indicated by a coefficient of -1; and zero indicates no association [55]. Time (hrs) and distance (nm) clearly show a significant positive correlation of 0.9654 from the heatmap, implying that the distance the ship covers also grows as the voyage time increases. Likewise, the time and LFO (gallons) show a positive association of 0.9123, meaning that longer travel distances are linked to greater fuel consumption. With a value of 0.8746, the association between the distance (nm) and LFO (gallons) is likewise strong; larger distances suggest higher fuel use. The plot visually highlights these associations using colour coding, which spans green (showing more correlation) to red (representing less correlation inside the range of the dataset).

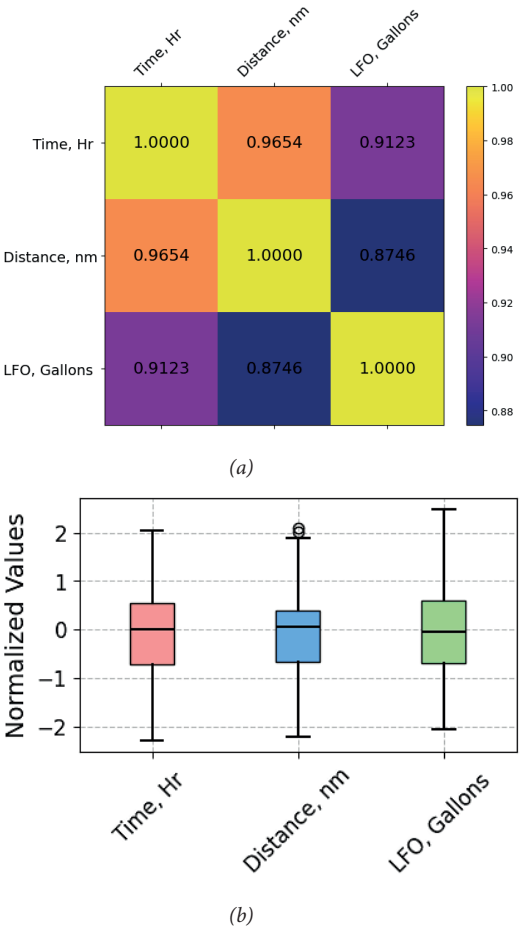
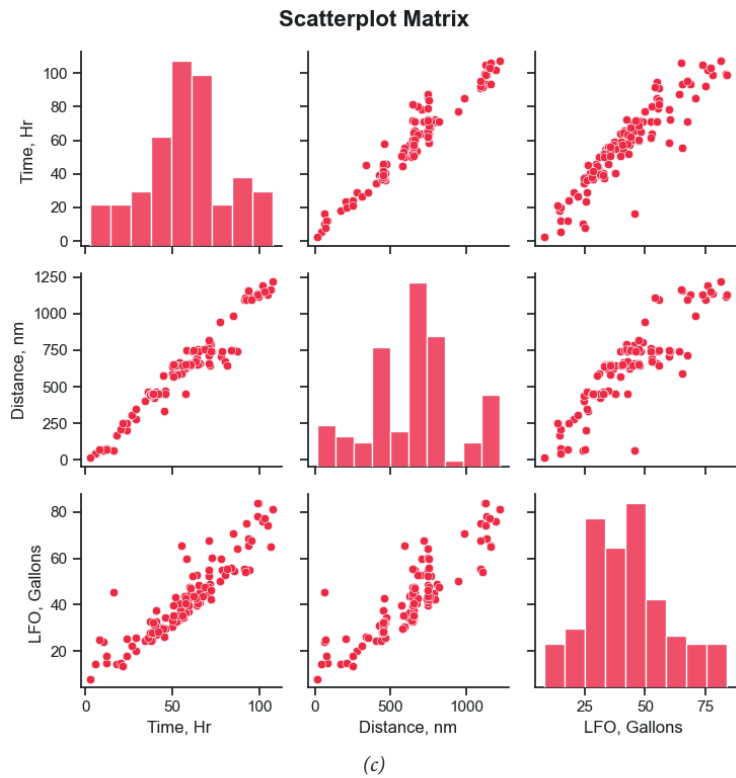


Fig. 1. Analysis of utilised data with (a) correlation heatmap; (b) box plot; (c) scatterplot matrix

Normalised values for ship fuel consumption data versus the time (hrs), distance (nm), and LFO (gallons) are shown in the box plot graph in Fig. 1(b). With medians centred within each box, each box shows the interquartile range (IQR). Whiskers show the data range, excluding outliers; there is one outlier in the distance variable. The whiskers indicate that the data are symmetrically distributed around the medians; time and the LFO usage point to balanced data distributions. There is one notable outlier, and the median of the distance variable is somewhat below the IQR centre. With few exceptions – except for the distance variable, which contains one obvious outlier – the data are generally well-distributed and free from anomalies. Fig. 1(c) depicts the scatterplot matrix for the data used in this study. Each variable's distribution is shown via the diagonal histograms. Time runs from around 0 to 100 hours; the peak falls around 50 to 60 hours. Most trips span 500 to 750 nautical miles; the distance spans 0 to over 1250 nautical miles. The LFO usage runs from 20 to 75 gallons, peaking at 50 to 60 gallons. Between all pairings of variables, including time and distance, time and LFO, and distance and LFO, the scatterplots off-diagonal demonstrate positive correlations. For example, ships running approximately 1000 nautical miles usually travel 60 to 100 hours and consume 50 to 75 gallons of LFO. This implies that the time spent rises proportionately as ships burn more fuel or cover more miles.

MODEL DEVELOPMENT AND EVALUATION

The forecasting models for ship fuel consumption were developed using five different ML approaches, i.e. LR, DT, RF, XGBoost, and AdaBoost. The data were randomly divided into training data (70%) and model testing data (30%). Open-source libraries were used in the Jupyter framework for model development. The developed models were utilised for forecasting the ship fuel consumption for the training and testing phases. The statistical evaluations of the model are listed in Table 2.

Tab. 2. Statistical evaluation of the developed model for ship fuel consumption

Model	Train MSE	Test MSE	Train R ²	Test R ²	Train MAPE	Test MAPE
Linear Regression	41.2	69.46	0.8307	0.8358	12.104	15.48
Decision Tree	0	56.80	1	0.8657	0	16.37
Random Forest	7.55	59.0	0.9689	0.8605	4.81	19.27
XGBoost	0.00031	64.1	0.9999	0.8483	0.034	15.99
AdaBoost	17.25	64.84	0.9291	0.8467	9.86	22.72

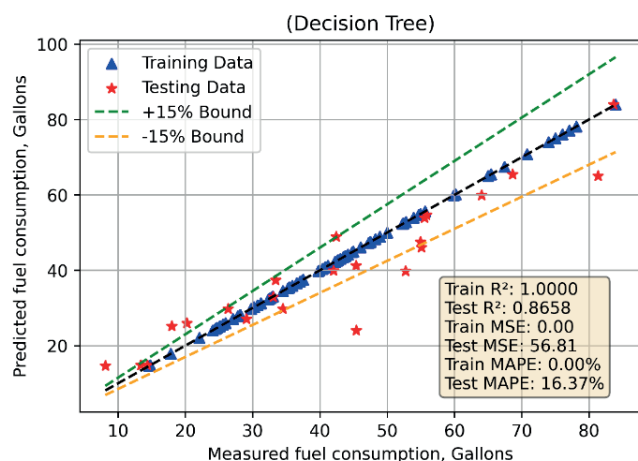
The prediction results in the case of the baseline LR-based model are compared in Fig. 2(a). It is observed that the model performed well, as most of the comparative data points are well within the ± 15 boundary lines. Based on the observed fuel usage, the graph shows a linear regression model

projecting ship fuel consumption. Blue triangles define the training data; red stars represent the testing data. Against the observed measured values, the regression line shows the expected values. Whereas the yellow dashed line suggests a 15% bound around the regression line, the green dashed line denotes a +15% bound. Important measures of the model performance are given: a strong fit is shown by the training R² of 0.8307 and the testing R² of 0.8359. Reflecting the accuracy of the model, the MAPE is 12.10% for training and 15.49% for testing; the MSE for training is 41.20, and for testing, it is 69.46. This implies that, with some variation in testing data relative to training data, the model consistently forecasts fuel use inside the specified limits. Based on observed fuel use, Fig. 2(b) shows a decision tree model projecting ship fuel consumption. Indicating almost no errors on the training set, the model obtains a perfect fit for the data with an R² of 1.0000, MSE of 0.00, and MAPE of 0.00%. Though much less than that for the training data, the R² value for the testing data is 0.8658, suggesting a high match. The DT model shows a few prediction errors but is still within tolerable ranges: the test MSE is 56.81 and the test MAPE is 16.37%. This implies that even with a somewhat greater error than the linear regression model, the decision tree model generalises effectively to the testing data, even if it fits the training data exactly.

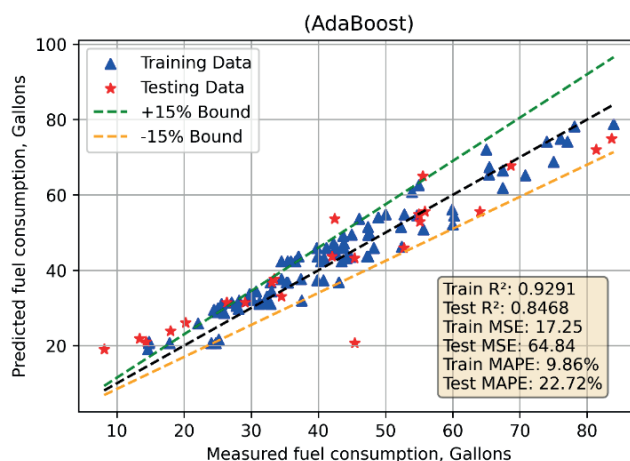
The model predictions from the RF model are depicted in Fig. 2(c). It shows robust predictions for both training and the model testing phase. In the training phase, the R² value was 0.9689, while it was 0.8605 in the model testing phase. This model obtained a low MSE value of 7.55 during training and an MSE of 59 during testing. The errors were also measured using the MAPE. The MSE values for the RF model were 4.81% and 19.27% during model training and model testing, respectively. This implies that while the performance of the RF-based model was superior to that of the LR model, it falls short of the DT-based ship fuel consumption model.



(a)

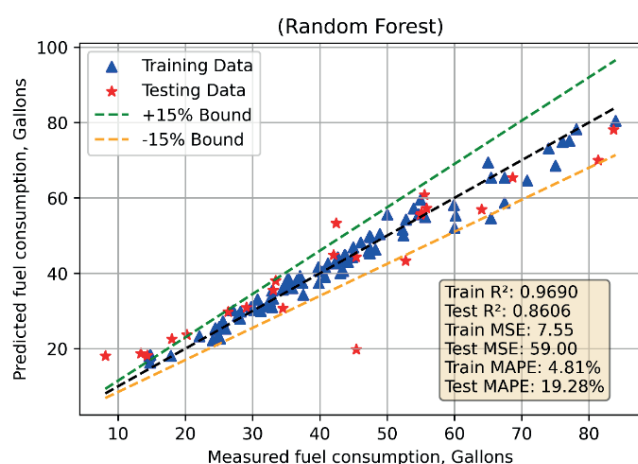


(b)

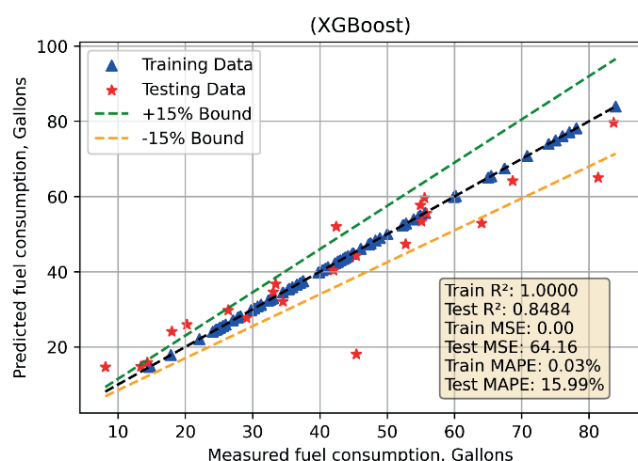


(e)

Fig. 2. Model performance and comparison for actual vs predicted fuel consumption in the case of (a) LR; (b) DT; (c) RF; (d) XGBoost; (e) AdaBoost



(c)



(d)

The findings given in Table 2 show the performance measures of the XGBoost-based ship fuel consumption model, therefore enabling the evaluation of both the model performance on unseen data and the training quality. In the case of the XGBoost-based model, a better match is shown by a reduced MSE, as it was quite low, reaching 0.00031. The model performance in the case of XGBoost is depicted in Fig. 2(d). Thus, this demonstrates that the model matches the training data flawlessly. The test MSE is 64.1, much higher, suggesting that the model does not perform as well on the testing data. R^2 gauges the percentage of dependent variable variation that may be expected given the independent variables. An R^2 around 1 means that the model nearly entirely explains the variation in the data. While the test R^2 is 0.8484, lower but still quite good, indicating that the model explains a significant percentage of the variation in the test data, the training R^2 is 0.9999, virtually perfect, showing a great fit to the training data. With a very low value of 0.03, the training MAPE shows a great fit to the training data. On the testing data, the test MAPE is 15.99, far higher, suggesting that the model's forecasts are less accurate. The low training MSE, high training R^2 , and low training MAPE point to the XGBoost model's high accuracy on the training data. In the case of AdaBoost, as depicted in Fig. 2(e), the results were not as good as for the XGBoost-based prediction framework. The training and test MSEs were 17.25 and 64.84, respectively. In the case of R^2 , the values attained were 0.9291 and 0.8467 during model training and testing, respectively. The MAPE values were 9.86 and 22.72 during model training and testing, respectively.

MODEL COMPARISON

The prediction models were compared using Taylor diagrams and violin plots. The Taylor diagram for the model training results is depicted in Fig. 3(a), while the Taylor diagram for model testing is illustrated in Fig. 3(b). In the case of model training, it can be observed that XGBoost and DT were the best-performing models, while in the case of the model testing phase, XGBoost was superior to the other four models. In another approach to model comparison, the hybrid representation of violin-box plots was employed in this study.

Fig. 4(a) shows hybrid violin-box plots for the model training phase, while Fig. 4(b) shows hybrid violin-box plots for model testing. These hybrid violin-box plots are used to compare the distribution and summary statistics of actual data (yellow) and predictions from many models, including LR (light green), XGBoost (salmon), ADB (AdaBoost) (violet), DT (cyan), and RF (green). Every plot features a box plot showing the median, quartiles, and any outliers, along with a violin plot displaying the kernel density estimate. During

model training, as depicted in Fig. 4(a), with a broader spread in the interquartile range compared to the actual data, the LR model has forecasts centred around a comparable median, showing some variability in its predictions. With a few additional outliers, the XGBoost model (salmon) likewise shows a distribution with a median near 50 and interquartile ranges similar to the actual data, but with small variations. Although AdaBoost shows a median around 50, its interquartile range is somewhat less than that of the real data, hence reducing the variability in predictions. With a median close to 50 but a greater interquartile range than the actual data, the DT model (cyan) shows higher forecast variability, suggesting that it represents the distribution of the real data very well. The RF model similarly has a median near 50, with interquartile ranges and an overall distribution generally comparable to the actual data. Based on its comparable median, interquartile range, and general distribution shape relative to the real data (yellow), the XGBoost model's predictions seem to be the most like the actual data, followed by random forest.

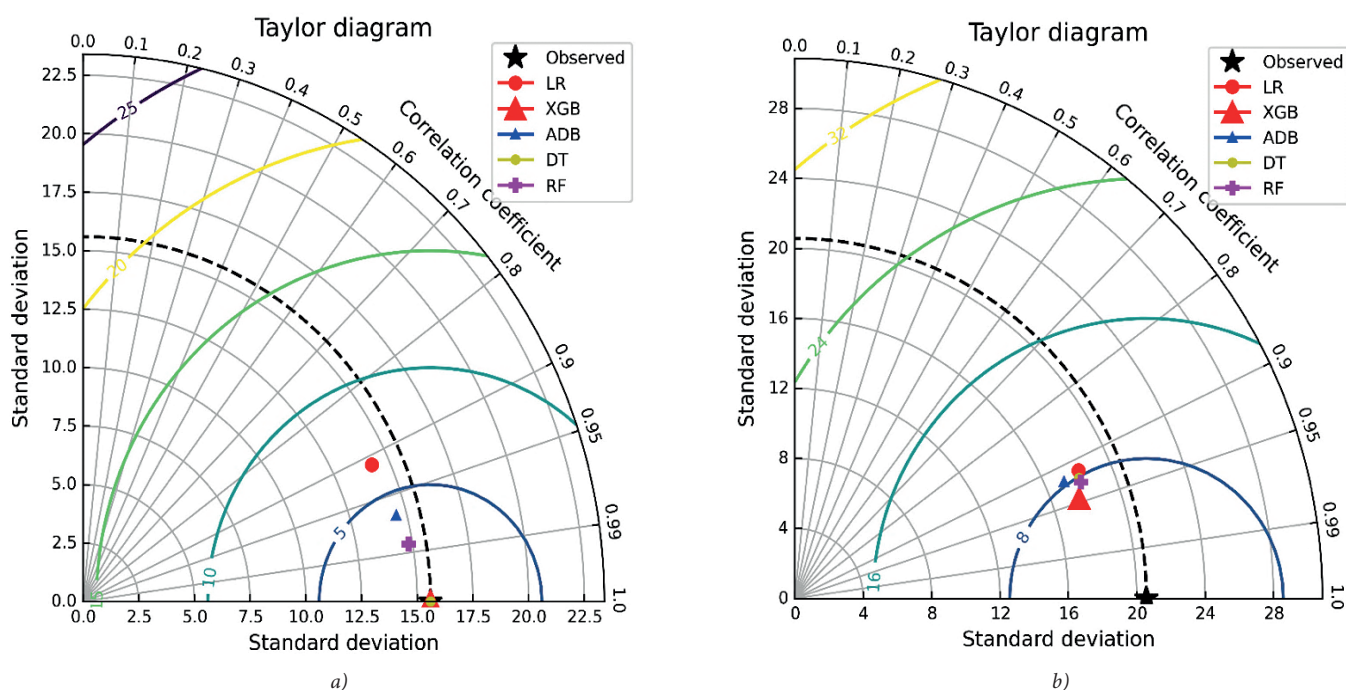
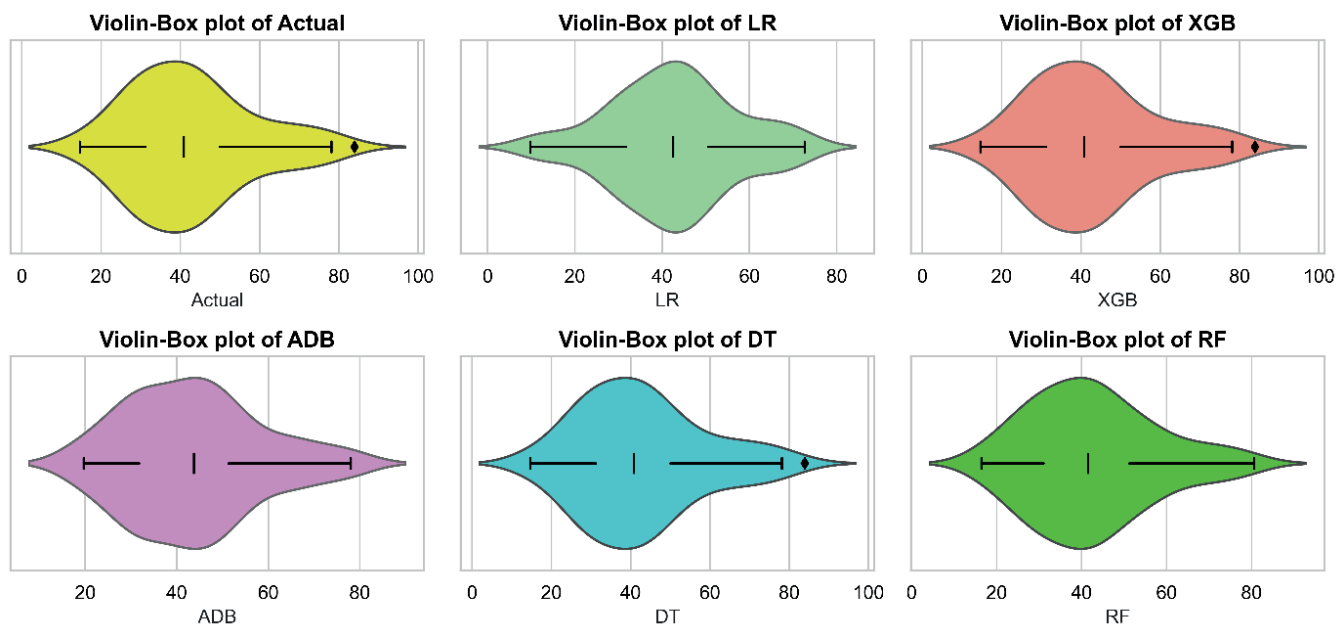
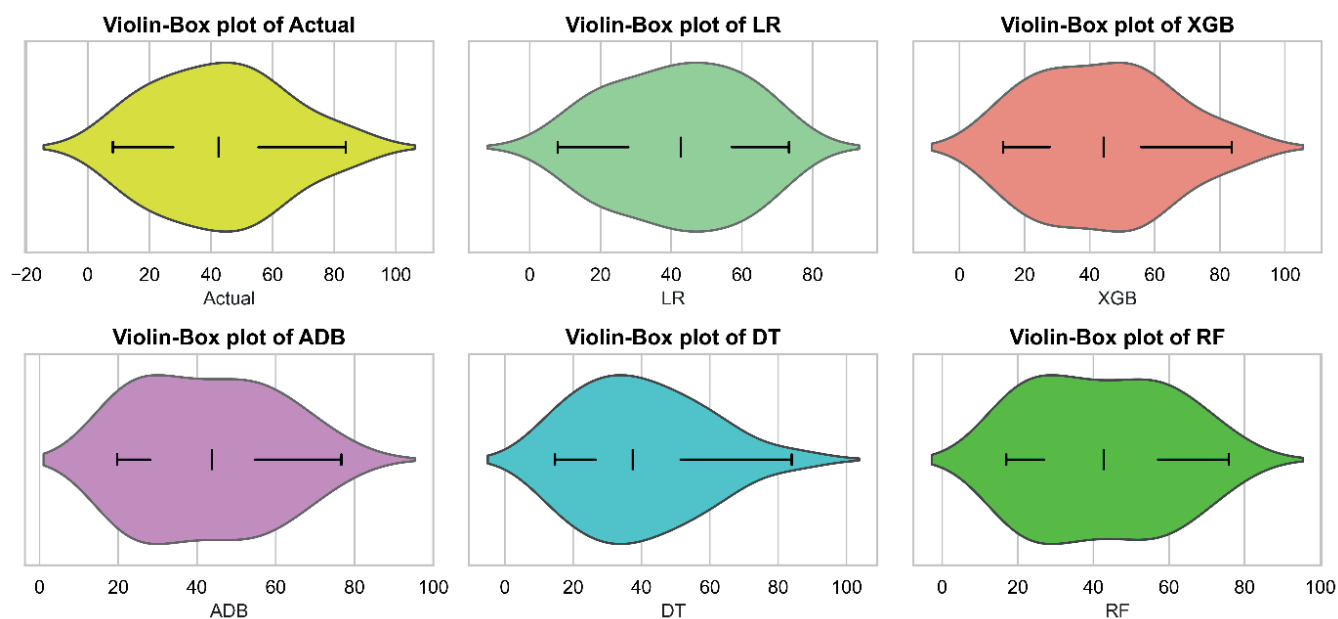


Fig. 3. Taylor diagram for model (a) training and (b) testing



(a)



(b)

Fig. 4. Violin plots for model (a) training and (b) testing

The hybrid violin-box plots in the case of model testing are depicted in Fig. 4(b), which shows how closely the model forecasts match the actual fuel usage. The LR model shows some deviations from the overall trend, as shown by the wider spread and density changes. Demonstrating accuracy and dependability, XGBoost works remarkably well and closely reflects the actual data distribution with little variance. Although it has somewhat more dispersion than XGBoost, the AdaBoost model also exhibits strong alignment. Although the DT model fits the training data exactly, overfitting is

indicated by a larger spread in the testing phase. Random forest demonstrates a strong performance since it balances preserving a narrow spread and capturing the trend. From the plots, the XGBoost and RF models track the real fuel usage trends with minimum spread and offer the most dependable predictions. Particularly, XGBoost is the best model since it has small prediction errors and a great fit, which makes it rather efficient for precise fuel consumption projection in ship operations.

EXPLAINABLE MACHINE LEARNING

Local interpretable model-agnostic explanations

The LIME map offers an understanding of the forecasts for ship fuel consumption provided by a machine learning model. The LIME plot, as depicted in Fig. 5, shows the importance of time (hours) and distance (nautical miles) for determining the outcome of the model. Represented in blue with a LIME weight of -9.9817, the distance ($\text{nm} \leq 450.88$) is the main contributing element. This shows that fuel consumption projections suffer significantly for lengths less than or equal to 450.88 nm. Shown in magenta with a LIME weight of -3.4631, the second most important variable is the time ($40.70 < \text{hr} \leq 58.05$). This spectrum of time also affects the expected fuel economy, albeit to less of a degree than distance. Through LIME, explainable ML approximates the complicated model locally with an interpretable model. This procedure perturbs the input data and tracks the changes in predictions to give weights to the characteristics depending on their relevance to the particular prediction. The graph clearly shows these weights, therefore enabling researchers to grasp and trust the model by clarifying which elements most affect its conclusions in a logical sense.

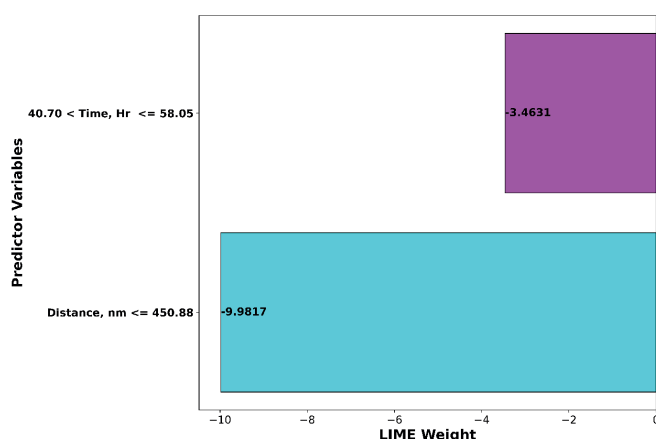


Fig. 5. LIME plots for ship fuel consumption

Shapley-based explanations

The SHapley Additive exPlanations (SHAP) feature importance plot is depicted in Fig. 6(a), while the SHAP summary graph is depicted in Fig. 6(b). The first plot (SHAP feature importance) is the respective importance of each feature. In this case, it is observed that the time that the ship sails is more important than the distance covered. The second plot (Fig. 6(b)) shows how input features affect a model's forecasts. Here, the x-axis shows the SHAP value, indicating the contribution of every feature to the output of the model. Positive SHAP values to the right of zero improve the expected result; negative SHAP values to the left have negative effects on the result. The y-axis lists the features; the main elements influencing the model are 'Time, Hr' and 'Distance, nm'. Every dot stands for a single data point with a feature value indicated by colour. Red dots indicate larger feature values;

blue dots indicate lower feature values. For instance, low values (blue) in the case of 'Time, Hr' generally have negative SHAP values, implying that shorter times lower the model's output (fuel consumption). Higher red values usually show positive SHAP values, meaning that longer times raise the expected result (higher fuel consumption). The majority of the SHAP values for 'Distance, nm' cluster around zero, showing that distance influences the output of the model less than time does. The dissemination of SHAP values also reveals the fluctuations in the way various feature values affect the predictions. In general, SHAP analysis demonstrated that time is a more important factor affecting the ship's fuel consumption compared to the distance travelled.

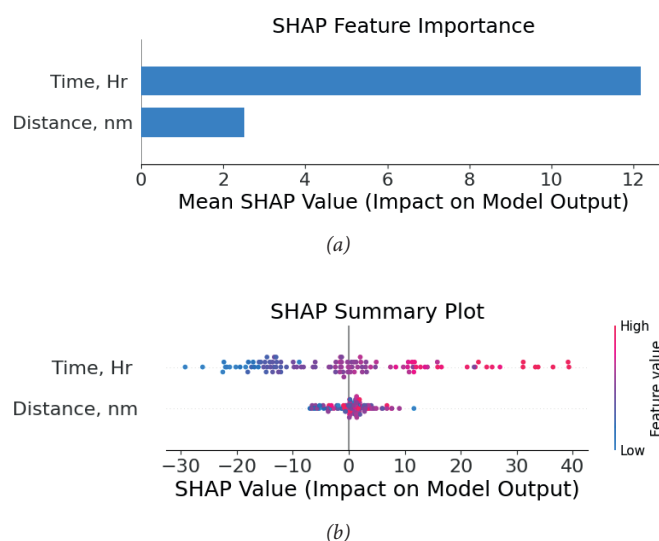


Fig. 6. SHAP plots for ship fuel consumption: (a) feature importance; (b) SHAP summary

Decision tree plot

The decision tree for the ship fuel consumption model is shown in Fig. 7. The decision tree diagram graphically shows the process of decision-making for the output variable. Starting with the root node (#0), which divides the data depending on 'Time, Hr ≤ 70.8 ', the tree decides whether the data move left (node #1), should the time be less than or equal to 70.8 hours, or right (node #16). Each subsequent node divides the data depending on various criteria; for example, node #1 considers 'Time, Hr ≤ 50.45 ', and node #2 considers 'Distance, nm ≤ 321.25 '. This process continues until the terminal nodes, where the tree generates a final prediction, are reached. Like node #4 or node #24, the leaf nodes mark the last results devoid of additional splits. Every leaf node shows the projected value for each sample, that is, the proportion of the dataset falling into that node. With higher nodes in the tree usually indicating more influential elements, the tree's decisions emphasise the need for time and distance in predicting the outcome. Obtaining predictions with this decision tree model helps one to clearly and interpretably understand how various elements, such as time and distance, affect the result. It presents a clear and orderly approach to generating forecasts and captures the interactions between

the input variables and the output. Pruning the tree to a maximum depth of 4 helps the model avoid overfitting, guaranteeing that it remains controllable and comprehensible.

In summary, the decision tree shows how the model was developed; it opens up the prediction model development process, making it a truly white box.

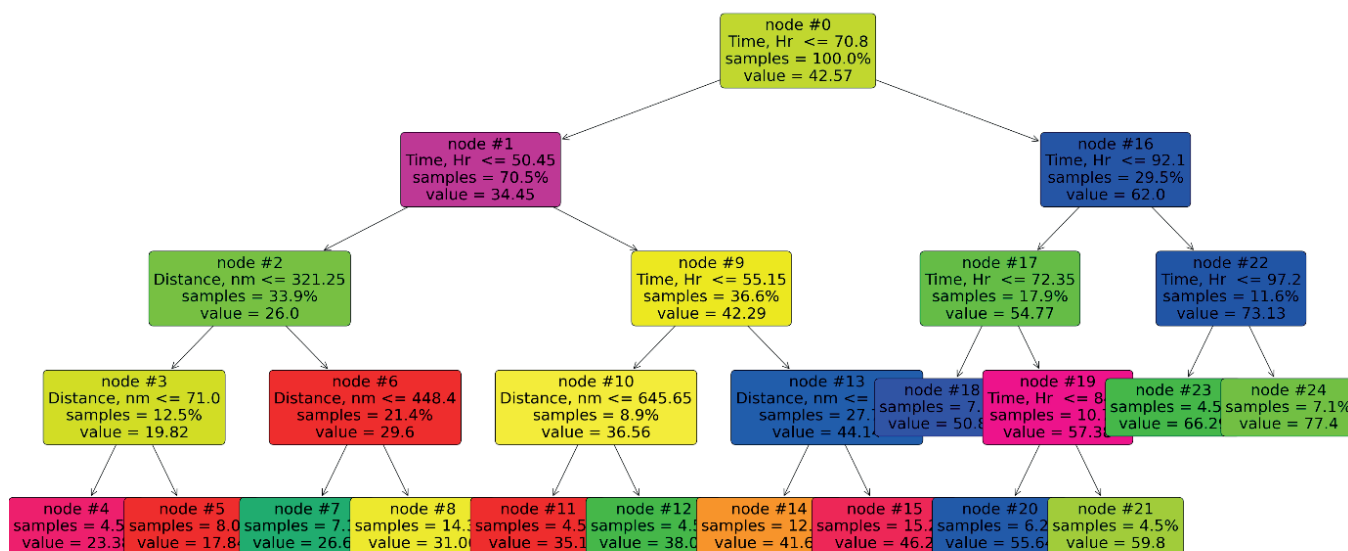


Fig. 7. Decision tree-based explanation of model development

CONCLUSION

In the present study, comprehensive forecasting models for ship fuel consumption were developed. Five different machine learning (ML) techniques, including linear regression (LR), decision tree (DT), random forest (RF), XGBoost, and AdaBoost, were employed. Furthermore, local interpretable model-agnostic explanations (LIME) were used to open up the black-box approach of machine learning. Using open-source modules in the Jupyter framework, the data were separated; 70% of the data were used for training and 30% for testing using the developed models. The models' capacity to forecast fuel use in both training and testing environments defined their evaluation. The following are the main outcomes of the study:

- The baseline LR model showed good performance. The test R^2 was 0.8359; the training R^2 was 0.8307. With mean squared errors (MSEs) of 41.20 and 69.46, respectively, the mean absolute percentage error (MAPE) for training was 12.10%, and for testing, it was 15.49%. This suggests a dependable forecast within given boundaries.
- In the case of the DT-based model, with an R^2 of 1.0000, an MSE of 0.00, and an MAPE of 0.00%, it attained an ideal fit on the training set. The R^2 value for testing data was 0.8658, the MSE was 56.81, and the MAPE was 16.37%, implying that the model generalises successfully despite occasional prediction errors.

- With a training R^2 value of 0.9689 and a testing R^2 of 0.8605, the RF model showed a strong prediction performance. The MSE values were 7.55 and 59; the MAPE values for training and testing, respectively, were 4.81% and 19.27%. Although not as high as the DT model, this shows the better performance of the RF model compared to the LR model.
- In the case of the XGBoost-based model, with an R^2 value of 0.9999, an MSE value of 0.00031, and an MAPE value of 0.034, it showed an outstanding training fit. With a test R^2 of 0.8484, an MSE of 64.1, and an MAPE of 15.99, its testing performance was somewhat lower, therefore emphasising its excellent training accuracy.
- With training and testing MSEs of 17.25 and 64.84, R^2 values of 0.9291 and 0.8467, and MAPE values of 9.86 and 22.72, respectively, AdaBoost performed quite moderately. The comparative investigation found that XGBoost and DT performed best during training; XGBoost exceeded the other models during testing.
- LIME provided an understanding of ship fuel consumption forecasting models. The distance (≤ 450.88 nm) was the primary contributing element. The time ($40.70 < \text{hr} \leq 58.05$) is the second most important factor that also negatively affects the forecasts.

REFERENCES

1. 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. <https://doi.org/10.62527/joiv.8.1.2581>.
2. Duc BV, Nguyen HP. A comprehensive review on big data-based potential applications in marine shipping management. *Int J Adv Sci Eng Inf Technol* 2021;11:1067–77.
3. UNCTAD. Review of Maritime Report 2021. 2021.
4. 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>.
5. Hoang AT, Foley AM, Nižetić S, Huang Z, Ong HC, Ölçer AI, et al. Energy-related approach for reduction of CO2 emissions: A critical strategy on the port-to-ship pathway. *J Clean Prod* 2022;355:131772. <https://doi.org/10.1016/j.jclepro.2022.131772>.
6. 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>.
7. Su W, Zhang Q, Liu Y. Event-triggered adaptive neural network trajectory tracking control for underactuated ships under uncertain disturbance. *Polish Marit Res* 2023;30:119–31. <https://doi.org/doi:10.2478/pomr-2023-0045>.
8. Sahren D, Harmsen J, Nesterova N, Bekdemir C, Van Kranenburg K, Andersson K, et al. Reducing sulphur emissions from ships - The impact of international regulation. *Int J Hydrogen Energy* 2020;12:32–5.
9. Le TT, Sharma P, Pham NDK, Le DTN, Le VV, Osman SM, et al. Development of comprehensive models for precise prognostics of ship fuel consumption. *J Mar Eng Technol* 2024;1–15. <https://doi.org/10.1080/20464177.2024.2372888>.
10. 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 Jatropa-based biodiesel. *J Mar Eng Technol* 2022;21:73–81. <https://doi.org/10.1080/20464177.2019.1595355>.
11. Nguyen VN, Chung N, Balaji GN, Rudzki K, Hoang AT. Internet of things-driven approach integrated with explainable machine learning models for ship fuel consumption prediction. *Alexandria Eng J* 2025;118:664–80. <https://doi.org/10.1016/j.aej.2025.01.067>.
12. 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>.
13. Mallouppas G, Yfantis EA. Decarbonization in shipping industry: A review of research, technology development, and innovation proposals. *J Mar Sci Eng* 2021;9:415. <https://doi.org/10.3390/jmse9040415>.
14. Herdzik J. Decarbonization of marine fuels—The future of shipping. *Energies* 2021;14:4311. <https://doi.org/10.3390/en14144311>.
15. 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>.
16. IMO. Fourth IMO GHG Study 2020 full report. *Int Marit Organ* 2021;6:524.
17. Chen Y, Sun B, Xie X, Li X, Li Y, Zhao Y. Short-term forecasting for ship fuel consumption based on deep learning. *Ocean Eng* 2024;301:117398. <https://doi.org/10.1016/j.oceaneng.2024.117398>.
18. Mantoju CD. Analysis of MARPOL implementation based on port state control statistics. *J Int Marit Safety, Environ Aff Shipp* 2021;5:132–45. <https://doi.org/10.1080/25725084.2021.1965281>.
19. Schinas O, Stefanakos CN. Selecting technologies towards compliance with MARPOL Annex VI: The perspective of operators. *Transp Res Part D Transp Environ* 2014;28:28–40. <https://doi.org/10.1016/j.trd.2013.12.006>.
20. Hussain N, Khan A, Shumaila, Memon S. Addressing marine pollution: An analysis of MARPOL 73/78 regulations and global implementation efforts. *J Soc Sci Rev* 2023;3:572–89. <https://doi.org/10.54183/jssrv3i1.193>.
21. Čampara L, Hasanspahić N, Vujičić S. Overview of MARPOL ANNEX VI regulations for prevention of air pollution from marine diesel engines. *SHS Web Conf* 2018;58:01004. <https://doi.org/10.1051/shsconf/20185801004>.
22. MEPC. MEPC 67 October 2014.
23. Chircop A. The IMO initial strategy for the reduction of GHGs from international shipping: A commentary. *Int J Mar Coast Law* 2019;34:482–512. <https://doi.org/10.1163/15718085-13431093>.

24. Joung T-H, Kang S-G, Lee J-K, Ahn J. The IMO initial strategy for reducing greenhouse gas (GHG) emissions, and its follow-up actions towards 2050. *J Int Marit Safety, Environ Aff Shipp* 2020;4:1–7. <https://doi.org/10.1080/25725084.2019.1707938>.
25. Nguyen VN, Rudzki K, Dzida M, Pham NDK, Pham MT, Nguyen PQP, et al. Understanding fuel saving and clean fuel strategies towards green maritime. *Polish Marit Res* 2023;30:146–64. <https://doi.org/10.2478/pomr-2023-0030>.
26. Ghaemi MH, Zeraatgar H. Impact of propeller emergence on hull, propeller, engine, and fuel consumption performance in regular head waves. *Polish Marit Res* 2022;29:56–76. <https://doi.org/10.2478/pomr-2022-0044>.
27. Korczewski Z. Energy and emission quality ranking of newly produced low-sulphur marine fuels. *Polish Marit Res* 2022;29:77–87. <https://doi.org/10.2478/pomr-2022-0045>.
28. Zhang M, Tsoulakos N, Kujala P, Hirdaris S. A deep learning method for the prediction of ship fuel consumption in real operational conditions. *Eng Appl Artif Intell* 2024;130:107425. <https://doi.org/10.1016/j.engappai.2023.107425>.
29. Agand P, Kennedy A, Harris T, Bae C, Chen M, Park EJ. Fuel consumption prediction for a passenger ferry using machine learning and in-service data: A comparative study. *Ocean Eng* 2023;284:115271. <https://doi.org/10.1016/j.oceaneng.2023.115271>.
30. Chen ZS, Lam JSL, Xiao Z. Prediction of harbour vessel fuel consumption based on machine learning approach. *Ocean Eng* 2023;278:114483. <https://doi.org/10.1016/j.oceaneng.2023.114483>.
31. Yan R, Wang S, Du Y. Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. *Transp Res Part E Logist Transp Rev* 2020;138:101930. <https://doi.org/10.1016/j.tre.2020.101930>.
32. 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>.
33. Shanmugasundar G, Vanitha M, Čep R, Kumar V, Kalita K, Ramachandran M. A comparative study of linear, random forest and AdaBoost regressions for modeling non-traditional machining. *Processes* 2021;9:2015. <https://doi.org/10.3390/pr9112015>.
34. Quirk TJ. Correlation and simple linear regression. *Excel 2016 Appl. Stat. High Sch. Students*, Cham: Springer International Publishing; 2018, pp. 107–52. https://doi.org/10.1007/978-3-319-89993-0_6.
35. Breiman L. Random forests. *Mach Learn* 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
36. Gnecco N, Terefe EM, Engelke S. Extremal random forests. *J Am Stat Assoc* 2024;1–24. <https://doi.org/10.1080/01621459.2023.2300522>.
37. 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>.
38. Sprenger M, Schemm S, Oechslein R, Jenkner J. Nowcasting Foehn wind events using the AdaBoost machine learning algorithm. *Weather Forecast* 2017;32:1079–99. <https://doi.org/10.1175/WAF-D-16-0208.1>.
39. Kyeremeh F, Zhi F, Yi Y, Gyamfi E, Nti IK. Solar PV power forecasting with a hybrid LSTM-AdaBoost ensemble. 2022 IEEE/IET Int. Util. Conf. Expo., IEEE; 2022, pp. 1–7. <https://doi.org/10.1109/IUCE55902.2022.10079424>.
40. Chen T, Guestrin C. XGBoost. *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, New York, NY, USA: ACM; 2016, pp. 785–94. <https://doi.org/10.1145/2939672.2939785>.
41. Darmawan H, Yuliana M, Hadi MZS. GRU and XGBoost performance with hyperparameter tuning using GridSearchCV and Bayesian optimization on an IoT-based weather prediction system. *Int J Adv Sci Eng Inf Technol* 2023;13:848–59. <https://doi.org/10.18517/ijaseit.13.3.18377>.
42. Abdi J, Hadavimoghaddam F, Hadipoor M, Hemmati-Sarapardeh A. Modeling of CO₂ adsorption capacity by porous metal organic frameworks using advanced decision tree-based models. *Sci Rep* 2021;11:24468. <https://doi.org/10.1038/s41598-021-04168-w>.
43. Sagi O, Rokach L. Approximating XGBoost with an interpretable decision tree. *Inf Sci (Ny)* 2021;572:522–42. <https://doi.org/10.1016/j.ins.2021.05.055>.
44. Jumin E, Basaruddin FB, Yusoff YBM, Latif SD, Ahmed AN. Solar radiation prediction using boosted decision tree regression model: A case study in Malaysia. *Environ Sci Pollut Res* 2021;28:26571–83. <https://doi.org/10.1007/s11356-021-12435-6>.
45. Sharma KV, Talpa Sai PHVS, Sharma P, Kanti PK, Bhramara P, Akilu S. Prognostic modeling of polydisperse SiO₂/Aqueous glycerol nanofluids' thermophysical profile using an explainable artificial intelligence (XAI) approach. *Eng Appl Artif Intell* 2023;126:106967. <https://doi.org/10.1016/j.engappai.2023.106967>.
46. Alruqi M, Sharma P, Algburi S, Khan MA, Alsubih M, Islam S. Biomass energy transformation: Harnessing the power of explainable AI to unlock the potential of ultimate

analysis data. *Environ Technol Innov* 2024;35:103652. <https://doi.org/10.1016/j.eti.2024.103652>.

47. de Myttenaere A, Golden B, Le Grand B, Rossi F. Mean absolute percentage error for regression models. *Neurocomputing* 2016;192:38–48. <https://doi.org/10.1016/j.neucom.2015.12.114>.
48. Elvidge S, Angling MJ, Nava B. On the use of modified Taylor diagrams to compare ionospheric assimilation models. *Radio Sci* 2014;49:737–45. <https://doi.org/10.1002/2014RS005435>.
49. Simão ML, Videiro PM, Silva PBA, de Freitas Assad LP, Sagrilo LVS. Application of Taylor diagram in the evaluation of joint environmental distributions' performances. *Mar Syst Ocean Technol* 2020;15:151–9. <https://doi.org/10.1007/s40868-020-00081-5>.
50. Aberasturi DT. Violin plot. *Wiley StatsRef Stat. Ref. Online*, Wiley; 2023, pp. 1–7. <https://doi.org/10.1002/9781118445112.stat08426>.
51. Molina E, Viale L, Vazquez P. How should we design violin plots? 2022 IEEE 4th Work. Vis. Guidel. Res. Des. Educ., IEEE; 2022, pp. 1–7. <https://doi.org/10.1109/VisGuides57787.2022.00006>.
52. Hintze JL, Nelson RD. Violin plots: A box plot-density trace synergism. *Am Stat* 1998;52:181. <https://doi.org/10.2307/2685478>.
53. Tanious R, Manolov R. Violin plots as visual tools in the meta-analysis of single-case experimental designs. *Methodology* 2022;18:221–38. <https://doi.org/10.5964/meth.9209>.
54. Vergura S. Supervision of the energy performance of multi-arrays PV plants by means of bi-monthly violin plots. *Renew Energy Power Qual J* 2022;20:625–8. <https://doi.org/10.24084/repqj20.384>.
55. Shin J, Yim I, Kwon S-B, Park S, Kim M, Cha Y. Evaluation of temperature effects on brake wear particles using clustered heatmaps. *Environ Eng Res* 2019;24:680–9. <https://doi.org/10.4491/eer.2018.385>.