International Journal of Advanced Smart Convergence Vol.13 No.3 66-79 (2024)
http://dx.doi.org/10.7236/IJASC.2024.13.3.66

IJASC 24-3-8

Development of a Carbon Emission Prediction Model for Bulk Carrier Based on EEDI Guidelines and Factor Interpretation Using SHAP

Hyunju Kim*, Byeongseok Yu**, Donghyun Kim***

**Senior Researcher, Department of Intelligent Convergence Research, Korea Marine Equipment Research Institute, Busan, Korea*
***, ***Professor, Department of Smart Machine Mobility Engineering, Pukyong National University, Busan, Korea*

hjkim94@komeri.re.kr, {youbs, dhkim8}@pknu.ac.kr

Abstract

The model developed in this study holds significant importance in predicting carbon emissions in maritime transport. By utilizing ship data and EEDI (Energy Efficiency Design Index) guidelines, the model presents a highly accurate prediction tool, providing a solid foundation for maximizing operational efficiency and effectively managing carbon emissions in ship operations. The model's accuracy was demonstrated by an R^2 score of 0.95 and a Mean Absolute Percentage Error (MAPE) of 1.4%. Through SHAP (SHapley Additive exPlanations) and Partial Dependence Plots (PDP), it was identified that Speed Over Ground and relative wind speed are the most significant variables, both showing a positive correlation with increased CO2 emissions. Additionally, environmental factors such as exceeding an average draft of 22(m), a Leeway over 5°, and a current angle exceeding 200° were found to increase emissions significantly. Specific ranges of wind and swell wave angles also notably affected emissions. Conversely, lower pitch, roll, and rudder angle were associated with reduced emissions, indicating that stable ship operation enhances efficiency.

Keywords: *Carbon Emission, Prediction Model, SHAP, Environmental Factor*

1. Introduction

The Internet of Things (IoT) technology is revolutionizing various industries, and the maritime sector is no exception. IoT enables real-time monitoring of vessel operations, positioning, weather analysis, fuel consumption, and engine status by collecting and analyzing data through sensors, devices, ports and other marine equipment[1]. These technological advancements significantly enhance the efficiency, safety, and sustainability of maritime operations by collecting and analyzing data through sensors, devices, ports, and other marine equipment. In particular, data collection through IoT sensors plays a crucial role in autonomous navigation, fuel efficiency improvement, and environmental monitoring. Recent studies have explored the

Manuscript Received: July. 9. 2024 / Revised: July. 16. 2024 / Accepted: July. 22. 2024
Corresponding Author: dhkim8@pknu.ac.kr
Tel: +82-051-629-7816, Fax: +82-051-629-6594
Author's affiliation: Professor, College of future convergence smart mechanical and mobility engineering, Korea

integration of blockchain, artificial intelligence (AI), and digital twin technologies to enhance the transparency and safety of vessel operations[2].

However, despite these technological advancements, the maritime industry remains a significant contributor to global carbon emissions. Despite the progress in monitoring and optimizing vessel operations, there is a clear need for more sophisticated models that can accurately predict carbon emissions based on a wide range of operational and environmental factors. IoT sensors and data analytics play a critical role in monitoring and optimizing fuel consumption patterns in real-time, contributing to the reduction of carbon emissions[3], [4]. Moreover, autonomous navigation technology contributes to reducing carbon emissions by maximizing fuel efficiency and minimizing human errors. These technological advancements are essential for improving the sustainability of the maritime industry.

At the 80th session of the Marine Environment Protection Committee (MEPC), the 'Net-Zero' strategy was approved, revising the target to reduce carbon emissions from the shipping sector by 50% to 100% by 2050 compared to 2008 levels[5]. To achieve a 100% reduction in carbon emissions compared to 2008 levels, IMO has implemented various regulations[6]. Among these regulations, the most important are the Energy Efficiency Design Index (EEDI)[7], the Energy Efficiency Operational Indicator (EEOI)[7], and the Carbon Intensity Indicator (CII)[8]. These environmental regulations aim to reduce the exhaust gases from ships. The CII represents the average emissions per transport work unit and has mandated annual carbon intensity reporting for international voyages of existing ships from 2023 onward. Ships that fail to meet regulatory standards face severe constraints, including potential operational cessation. Compliance with the CII requires the accurate real-time prediction of carbon emissions. The ability to accurately monitor and optimize carbon emissions is considered essential for meeting stringent CII requirements[9].

In this paper, we develop a carbon emission prediction model utilizing factor values derived from EEDI standards and use SHAP (SHapley Additive exPlanations) to interpret the influence of environmental factors on the model's predictions. This approach not only improves compliance with environmental regulations but also supports the industry's broader efforts to achieve sustainability. Chapter 2 outlines the research methodology for constructing the prediction model. Chapter 3 presents the analysis of the model's results, and Chapter 4 concludes the study with a discussion of the findings and suggestions for future research.

2. Research Methods

2.1 Target ship Description

A large bulk carrier with a gross tonnage of 160,290 tons and a length overall of 333 meters was utilized, as shown in Table 1. The ship has a beam of 60 meters, a depth of 29 meters, and it operates at a nominal continuous rating of 59 rpm. The ship is designed to achieve a speed of 14.5 knots with a summer draft of 22 meters. This research utilized data from the main engine sensors, which were collected every six months. Additionally, the dataset comprises wind direction and wind speed data from sensors installed on the ship, positional data of the ship, and external meteorological information such as wave height and current conditions. While a total of 190 features were initially collected, this study focuses on 21 key features. The final dataset consists of 24,854 rows of data corresponding to the selected features. The ship operates as a regular liner on a round-trip route between South Korea and South America, as shown in Figure 1.

Table 1. Ship Specifications

Specification	Measurement	Unit
Length Overall	333	m
Beam	60	m
Depth	29	m
Gross Tonnage	160,290	ton
Summer Draft	*22*	m
Designed Speed	*14.5*	knots
Nominal Continuous Rating	*59*	rpm

Figure 1. Ship route

The study utilized a diverse set of approximately 190 features related to the vessel's operation. However, having too many features can lead to overfitting, making it crucial to carefully select features to include in the model. Among the navigation and engine data features are those manipulated by navigators and engineers during vessel operations, as well as the resulting output features. Representative features include ground speed, heading, and rudder angle, which are adjusted to determine sailing speed and route. On the other hand, features such as RPM and Power are related to engine output resulting from the manipulated features, with higher values expected to correspond to increased fuel consumption. Therefore, in this study, engine output features were excluded from the candidate features for the model due to their direct correlation with fuel consumption. These features are detailed in Table 2.

Table 2. Original Feature List

Feature Name	Unit	Description
ME1_FOC	kg/h	No. 1. Main Engine Fuel Oil Consumption
SPEED_VG	knots	Speed Over Ground

DRAFT_FORE	m	Draft at the front (bow) of the ship.
DRAFT_AFT	m	Draft at the rear (stern) of the ship.
DRAFT_MID_PORT	m	Draft at the midsection, port (left) side.
DRAFT_MID_STBD	m	Draft at the midsection, starboard (right) side.
SHIP_HEADING	°	Direction the bow is pointing.
COURSE_OVER_GROUND	°	Direction of the ship's movement as measured by GPS.
RUDDER_ANGLE	°	Rudder angle
SWELL_WAVE_DIRECTION	°	Direction from which long-period ocean swells originate.
WIND_WAVE_DIRECTION	°	Direction from which wind-generated waves come.
CURRENT_DIRECTION	°	Compass direction of ocean current flow.
REL_WIND_DIR	°	Angle of wind relative to the ship's heading..
CURRENT_SPEED	m/s	Speed of ocean current, influenced by gravity, wind, and water density..
REL_WIND_SPEED	m/s	Wind speed relative to the ship's speed.
SWELL_WAVE_PERIOD	s	Time interval between successive swells.
WIND_WAVE_PERIOD	s	Time interval between successive phases of a wind wave.
SWELL_WAVE_HEIGHT	m	Height of long-period waves, unaffected by local winds.
WIND_WAVE_HEIGHT	m	Height of wind-driven waves on the sea surface.
ROLL	°	Ship's side-to-side (lateral) movement.
PITCH	°	Ship's up-and-down (longitudinal) movement.

2.2 Data Filtering and Cleaning

When operating a ship, most voyages are planned in advance along predetermined routes and navigation plans. Additionally, when the ship operates at a certain speed, the engine functions more stable[11], [12], [13]. Therefore, filtering the data based on specific criteria can enhance the accuracy of the model development. Under low-speed or abnormal operating conditions, the engine's operational efficiency may vary, which can negatively affect the model's results. Consequently, in this study, we filtered the data to include only instances where the ship's speed was above 10 knots and the RPM was above 50, considering the ship's designed speed and Normal Continuous Rating (NCR). After performing data filtering, the data was reduced from 24,855 rows to 16,584 rows, resulting in the removal of 8,271 rows. This means that approximately 33.3% of the data was discarded.

Sensor data errors and missing values on ships can arise due to various factors such as sensor contamination from marine growth, inadequate calibration, communication issues, and data loss[10].

In the target ship used for this study, errors and missing data were recorded as -9999 or marked as "error". Consequently, during preprocessing, rows containing these errors or missing values were removed to ensure data quality and accuracy in subsequent analyses. Through this process, the final dataset was reduced from 16,854 rows to 16,542 rows after removing 42 rows with errors. Of the 16,542 rows, 80%, or 13,233 rows, were used for training. There are methods to replace missing or abnormal data with the mean or median values. However, in the case of ships, due to the critical importance of accurately calculating the CII regulatory rating, which directly impacts operational restrictions, it was decided to remove the data rather than impute it, following expert recommendations.

2.3 Data Featuring

Various feature engineering techniques were applied to ship operation data to predict CO_2 emissions. This process involved extracting meaningful information from the raw data and generating variables that enhance the model's predictive performance. The features developed in this study represent the physical state and environmental conditions of the ship and were utilized as input variables for the CO2 emission prediction model. Additionally, these feature engineering techniques were used to apply SHAP analysis to understand the impact of environmental factors on CO2 emissions.

The average draft (AvgDraft) indicates the ship's center of gravity, while the leeway (Leeway) serves as a indicator of course stability. The current angle (Current Angle), wind wave angle (Wind Wave Angle), and swell wave angle (Swell Wave Angle) quantify the impact of ocean currents, wind waves, and swell waves on ship navigation, respectively. The CO2 emission (CO2 Emission), which is the target variable predicted in this study, is derived directly from the ship's fuel consumption data.

Each of the generated variables is described by equations (1) through (5). The average draft was determined by calculating the mean of the draft values measured at the front and rear, as well as the left and right sides of the ship. This average draft serves as an indicator of the ship's submerged depth, representing how deeply the ship sits in the water. The calculation was performed using Equation (1).

$$Avg\ Draft = \frac{(D_{fore}+D_{aft}+D_{mid-port}+D_{mid-stbd})}{4} \tag{1}$$

Leeway in Equation 2 is a metric used to evaluate the directional stability of a ship by calculating the angle difference between the ship's actual heading and the planned course. In this calculation, the planned course (Course Over Ground, COG) is subtracted from the ship's actual heading (Ship Heading, SHD). The reason for adding 180 degrees to this difference is to avoid negative angles and to adjust the result within the range of 0 to 360 degrees. Afterward, a modulo operation is applied to ensure the angle remains within the range of 0 to 360 degree. Finally, 180 degrees are subtracted from the calculated value to derive the final Leeway, which falls within the range of -180 to 180 degrees. This value quantitatively indicates how much the ship has deviated from its planned course and plays an important role in assessing the ship’s directional stability.

$$Leeway = \left(\left(\theta_{SHD} - \theta_{COG} + 180^{\circ}\right) mod\ 360\right) - 180^{\circ} \tag{2}$$

Equation 3 describes the Current Angle, a metric that indicates the angular difference between the ship's heading and the direction of the ocean current, playing a crucial role in assessing how much the ship is influenced by the current during navigation. This angle is calculated by determining the difference between the Current Direction and the ship's actual heading (Ship Heading). First, the ship's heading is subtracted from the current's direction, and then 360 degrees are added to prevent the angle from becoming negative. The result is then adjusted within the range of 0 to 360 degrees by applying a modular operation with 360. The Current Angle thus ranges from 0 to 360 degrees, quantitatively indicating the impact of the current on the ship's course. For example, if the Current Angle is close to 0 degrees, the current aligns with the ship's heading, while an angle close to 180 degrees suggests that the current flows in the opposite direction of the ship's heading.

$$Current\ Angle = \left(\left(\theta_{current} - \theta_{SHD} + 360^{\circ}\right) mod\ 360\right) \tag{3}$$

As described in Equations 4 and 5, the Wind Wave Angle and Swell Wave Angle were calculated using the same feature engineering approach used for the Current Angle. Wind Wave Angle ranges from 0 to 360 degrees and represents the angular difference between the ship's heading and the direction of wind-driven waves. For instance, a Wind Wave Angle close to 0 degrees indicates that the wind waves are coming from a direction aligned with the ship's heading, while a Wind Wave Angle near 180 degrees suggests that the wind waves are coming from the opposite direction of the ship's heading. Swell Wave Angle also ranges from 0 to 360 degrees and reflects the angular difference between the ship's heading and the direction of long-period swell waves.

$$Wind\ Wave\ Angle = \left(\left(\theta_{wind\_wave} - \theta_{SHD} + 360^{\circ}\right) mod\ 360\right) \tag{4}$$

A Swell Wave Angle close to 0 degrees signifies that the swell waves are aligned with the ship's heading, whereas a Swell Wave Angle near 180 degrees indicates that the swell waves are coming from the opposite direction.

$$Swell\ Wave\ Angle = \left(\left(\theta_{swell\_wave} - \theta_{SHD} + 360^{\circ}\right) mod\ 360\right) \tag{5}$$

CO2 emissions were calculated based on the main engine fuel oil consumption (ME1_FOC). The vessel used in the study was a diesel ship, and upon reviewing the daily log, it was observed that there had been a high usage of HFO (Heavy Fuel Oil). Therefore, the carbon factor value for HFO suggested in the EEDI guidelines was utilized. According to the EEDI guidelines, the carbon factor for HFO is 3.114 g of CO2 per gram of fuel oil. In this study, fuel oil consumption was measured in kg per hour. Thus, the units of fuel oil consumption were aligned with the formula, as shown in Equation 6, and the necessary feature transformation was performed accordingly.

$$CO2\ Emission = \quad FOC_{ME1} * 3114 \tag{6}$$

The list of transformed features can be found in Table 3. Through this process of feature transformation, new features are created to quantify the differences between external environmental factors, such as wind and currents, and the ship's navigational direction. The aim is to gain a more detailed understanding through SHAP of how these influences impact ship carbon emissions.

Table 3. Feature Transformation List

Feature Name	Unit	Description
Avg Draft	m	The average draft at the bow, stern, port side, and starboard side.
Leeway	°	Difference in angles between Course Over Ground and Ship Heading
Current Angle	°	Difference in angles between the current direction and Ship Heading
Wind Wave Angle	°	Difference in angles between the direction of wind waves and Ship Heading
Swell Wave Angle	°	Difference in angles between the direction of swell waves and Ship Heading
CO2 emission	kg/h	Carbon emissions from the No.1 main engine

2.4 CO2 Emission Prediction

Since the carbon emissions of a ship are determined by various operational conditions, a model was developed that includes a range of environmental features. Additionally, given that the ship's large dataset consists mostly of numerical and categorical data, The XGBoost (eXtreme Gradient Boosting) was employed for model development due to its suitability for such characteristics. The XGBoost (eXtreme Gradient Boosting) model is an extension and improvement of the Gradient Boosting algorithm, which combines multiple weak learners, specifically decision trees, to create a more powerful ensemble model than a single decision tree[14]. It offers better predictive performance than a single model and can mitigate overfitting due to its sequential learning process[14], [15]. Furthermore, XGBoost includes overfitting prevention techniques such as L1 and L2 regularization, ensuring stable performance even with large datasets. While other regression models can also handle numerical and categorical data, XGBoost was chosen for this study due to its ability to effectively manage complex data interactions, provide fast computation, and deliver high predictive accuracy. The XGBoost model was optimized for the ship's dataset by performing hyperparameter tuning using GridSearchCV. Key hyperparameters included n_estimators (the number of trees in the model), max_depth (the maximum depth of each tree), and learning_rate (step size shrinkage used in updates to prevent overfitting). The ranges explored during the optimization process were n_estimators: 50, 100, 150, 200, 250; max_depth: 5, 10, 25, 50; and learning_rate: 0.1, 0.01, 0.001. This comprehensive search enabled the selection of the optimal model by identifying the best combination of hyperparameters to improve predictive performance.

Additionally, SHAP was utilized to interpret the model results. SHAP is a model explanation technique widely used in the field of XAI (Explainable Artificial Intelligence) to interpret the predictions of machine learning models. This method is based on the Shapley value from game theory, which quantitatively evaluates the contribution of each input variable to the predictions. SHAP values calculate the importance of each variable and numerically explain the positive or negative impact that changes in these variables have on the prediction outcomes[16]. Through this, SHAP provides a powerful tool for interpreting not only individual predictions but also the overall behavior of the model. In particular, SHAP considers the interactions between variables when explaining prediction results, allowing for a more sophisticated understanding of variable importance in complex models. This methodology enhances the interpretability of models and increases their practical applicability in real-world scenarios[17].

3. Result and Discussion

3.1 CO2 Emission Prediction Result

As shown in Table 4, this model demonstrates exceptionally high accuracy on the training data. The mean Absolute Error (MAE) is 19,401, which is a very small error considering the data points are in the range of millions. The mean absolute percentage error (MAPE) is 0.33%, indicating that the model's predictions are nearly identical to the actual values. The R^2 score of 0.99 suggests that the model explains 99.9% of the variance in the training data. On the test data, the MAE is 89,917, which is higher than in the training data, but still relatively small given the scale of the data. The MAPE is 1.4%, showing that the model maintains a high level of accuracy even on unseen data. The R^2 score of 0.95 indicates that the model accounts for 95.5% of the variance in the test data. These results can be further confirmed through the visualization in Figure 2. The predicted values, shown in blue, accurately capture the variability of the actual values, depicted in red, across each data point.

Table 4. Prediction Result

	Train	Test
R-squared	0.99	0.95
MAE	19,401	89,917

MAPE	0.0033	0.014

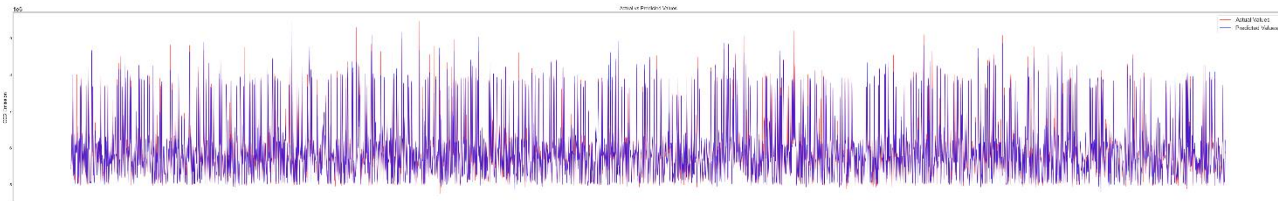

Figure 2. Actual value and Predicted Value

3.2 SHAP Result

Figure 3 presents the average impact of each feature on the model's predictions using a SHAP value bar chart. The SHAP values quantify the contribution of each feature to the model's prediction, with higher values signifying greater influence. The features are ordered by descending importance, offering insights into which factors most significantly affect the model's output.

At the top of the chart, SPEED_VG emerges as the most influential feature, with an average SHAP value of +195,682.07. This substantial value indicates that the vessel's speed is a critical determinant of CO2 emissions. Higher speeds typically result in increased fuel consumption, directly leading to higher CO2 emissions, which explains the dominant role of this feature in the model's predictions. Following SPEED_VG, REL_WIND_SPEED ranks as the second most important feature, with a SHAP value of +127,031.26. This feature reflects the relative wind speed, which has a significant impact on the ship's fuel efficiency. High wind speeds, particularly when opposing the direction of travel, necessitate greater power to maintain speed, thereby increasing fuel consumption and, consequently, CO2 emissions. The model effectively captures this relationship, as evidenced by the substantial SHAP value.

The next two features, DRAFT_FORE and PITCH, exhibit SHAP values of +117,319.68 and +100,192.63, respectively. DRAFT_FORE refers to the draft at the ship's bow, while PITCH refers to the angle of the ship's pitch, which is the tilt from front to back. Both features are crucial because they influence the vessel's hydrodynamic efficiency. A deeper draft and significant pitch increase water resistance, requiring more energy (and fuel) to maintain speed, thereby elevating CO2 emissions. The model captures these dynamics, assigning high SHAP values to these features.

DRAFT_MID_STBD, with a SHAP value of +81,133.05, also plays a significant role in the model's predictions. This feature represents the draft at the midship starboard (right side of the ship). Similar to DRAFT_FORE, it affects the ship's stability and water resistance, impacting fuel consumption and emissions. SWELL_WAVE_HEIGHT and AvgDraft have SHAP values of +60,616.02 and +52,284.69, respectively. SWELL_WAVE_HEIGHT measures the height of ocean swells, which can affect the ship's motion and stabilitythe ship's motion and stability. Higher swells increase the ship's resistance in the water, necessitating more power to maintain speed. AvgDraft, which represents the average draft of the ship, similarly affects the ship's submersion and the resistance it encounters. Both features contribute meaningfully to the model's predictions, as reflected in their SHAP values.

ROLL and WIND_WAVE_HEIGHT are also important, with SHAP values of +43,107.18 and +35,067.44, respectively. ROLL pertains to the ship's side-to-side tilting motion, which can affect fuel efficiency by increasing drag. WIND_WAVE_HEIGHT, which measures the height of wind-generated waves, also influences the ship's passage through water, further impacting fuel consumption and CO2 emissions.

Other features including DRAFT_AFT, WIND_WAVE_PERIOD, and SHIP_HEADING also demonstrate notable contributions to the model's predictions, although their SHAP values are relatively lower. These features still play roles in determining the ship's fuel efficiency and CO2 emissions, albeit to a lesser extent than the top-ranked features.

In conclusion, the SHAP value bar chart in Figure 3 provides a clear visualization of the most critical features influencing CO2 emissions. SPEED_VG and REL_WIND_SPEED are the primary factors, followed by various draft measurements and wave-related features. This analysis underscores how different aspects of ship operation and environmental conditions combine to influence CO2 emissions, with the model effectively capturing these relationships through its SHAP values.

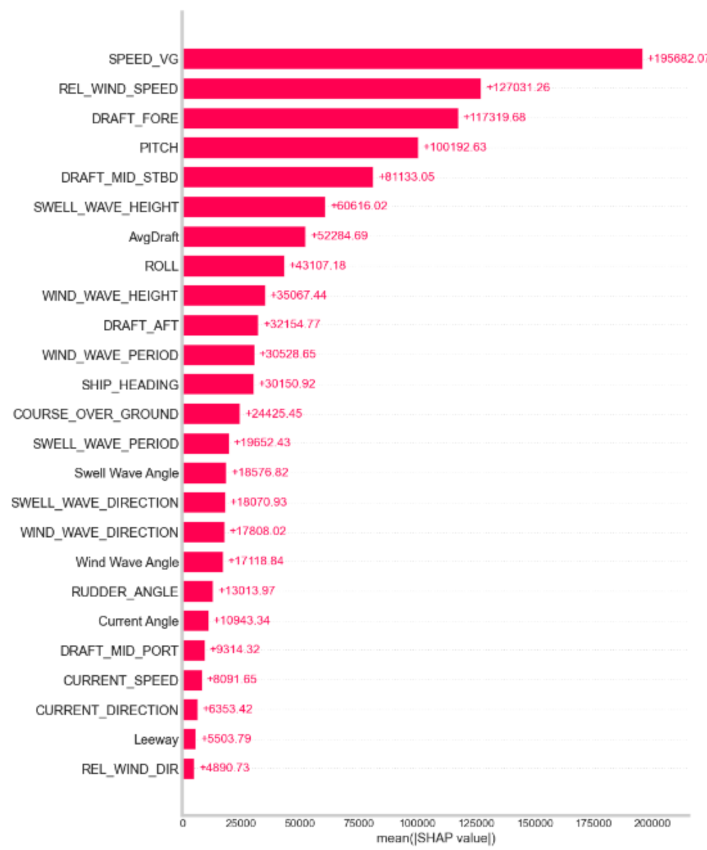

Figure 3. Mean SHAP Values Indicating Feature Importance in Model Predictions

Figure 4 provides a more detailed understanding of whether these features contribute to an increase or decrease in the model's prediction values. SPEED_VG, which primarily exhibits a positive relationship, is the most influential variable in the model's predictions. As shown in the graph, higher speeds (indicated in red) are predominantly positioned to the right, contributing to an increase in predicted CO2 emissions. This occurs because of higher ship speeds leading to increased fuel consumption, which in turn raises CO2 emissions. REL_WIND_SPEED also plays a significant role in the model's predictions, generally displaying a positive relationship. Higher relative wind speeds tend to increase the predicted values. This can be interpreted as wind speed and direction increasing the ship's resistance, thereby requiring more fuel and consequently resulting in higher CO2 emissions. SWELL_WAVE_HEIGHT and WIND_WAVE_HEIGHT also exhibit a positive relationship with the model's predicted values, suggesting that as these variables increase, the predicted CO2 emissions also rise. This is likely due to these conditions increasing the ship's resistance to motion, thereby reducing its efficiency.

PITCH represents the pitch angle of the ship, with lower PITCH values (indicated by blue points) tending to decrease the model's predicted values. This negative relationship suggests that a smaller pitch angle allows the ship to navigate more stably, leading to reduced fuel consumption and consequently lower CO2 emissions. Similarly, ROLL, which indicates the degree of the ship's side-to-side tilting, shows that lower ROLL values (blue points) are associated with decreased predicted values. Lower DRAFT_FORE and DRAFT_MID_STBD values also contribute to a decrease in CO2 emissions, as shallower drafts reduce the ship's resistance, leading to lower fuel consumption and subsequently reduced CO2 emissions. WIND_WAVE_PERIOD represents the period of waves generated by the wind. Lower wave periods (blue points) tend to decrease the model's predicted values, reflecting the tendency for shorter wave periods to have a lesser impact on fuel consumption. RUDDER_ANGLE (the angle of the ship's rudder) also exhibits a negative relationship. Lower rudder angles (blue points) correlate with lower predicted values, indicating that maintaining a straight course reduces fuel consumption and consequently lowers CO2 emissions.

Features such as Avg Draft, Leeway, Current Angle, Wind Wave Angle, and Swell Wave Angle may appear to have lower importance, but they can play critical roles in specific situations and enhance the model's interpretability. For instance, these variables are useful for accurately assessing the impact of extreme weather conditions or specific routes on fuel consumption and carbon emissions. Including these features allows for more accurate predictions by considering interactions between various variables. Additionally, by examining specific values through PDP plots, these features can be used as important indicators for ship operation management.

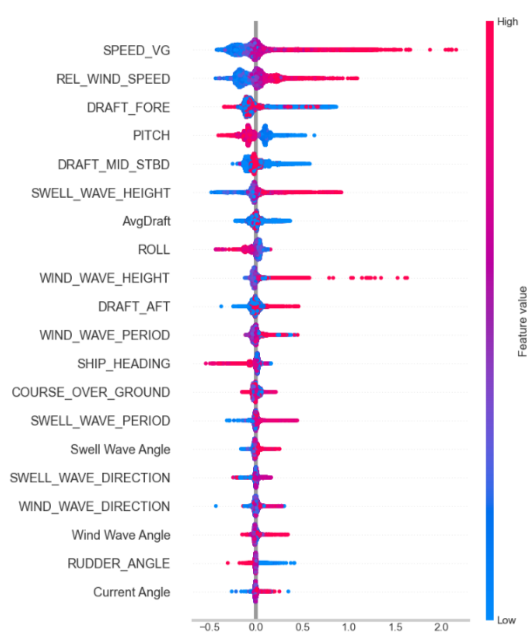

Figure 4. SHAP Values for Feature Impact on Model Output

To further understand the impact of external environmental factors on a ship's carbon emissions, Partial Dependence Plot(PDP) was generated using the transformed features shown in Figures 5~9. These transformed features reflect the differences between each environmental factor and the ship's inherent heading, focusing on how these differences influence the model's predictions. PDP provides a visual explanation of how a black-box model responds to specific variables. They help to identify which variables the model considers most important for making predictions, and how changes in those variables affect the predicted outcomes. Additionally, PDP offers concrete numerical insights that can be combined with domain knowledge to achieve a deeper interpretation of the model's predictions.

According to Figure 5, when the AvgDraft is approximately 16 or below, the model's predictions remain relatively stable. However, as the AvgDraft increases to around 22 or higher, the predicted values rise sharply. This indicates that a deeper draft causes the ship to submerge more, increasing resistance, which in turn significantly elevates fuel consumption and CO2 emissions. Furthermore, this aligns with the ship's originally designed summer draft of 22, suggesting that surpassing this design limit leads to a notable decrease in the ship's efficiency, thereby causing a sharp increase in the model's predicted values.

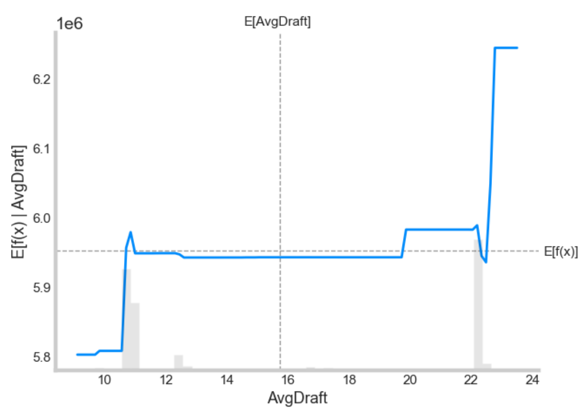

Figure 5. PDP Plot of AvgDraft

According to Figure 6, when the Leeway is close to 0, the model's predictions remain stable. However, when the Leeway exceeds 5°, the predicted values increase sharply. This suggests that when the ship deviates from its intended course due to various environmental factors (such as wind, waves, etc.), the resistance increases as the ship attempts to maintain its course, likely resulting in higher fuel consumption and CO2 emissions.

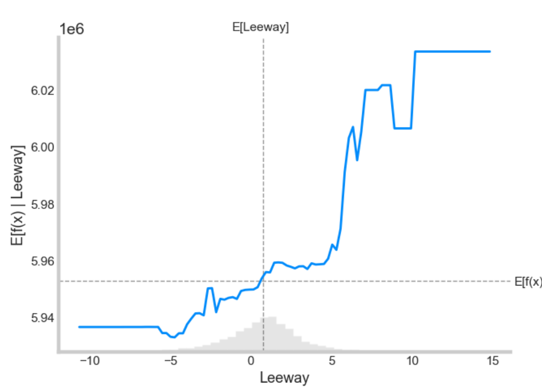

Figure 6. PDP Plot of Leeway

According to Figure 7, when the current angle exceeds approximately 200°, the model's predictions tend to increase sharply. This suggests that when the current is significantly misaligned with the ship's heading, the ship encounters greater resistance, which can lead to increased fuel consumption and CO2 emissions.

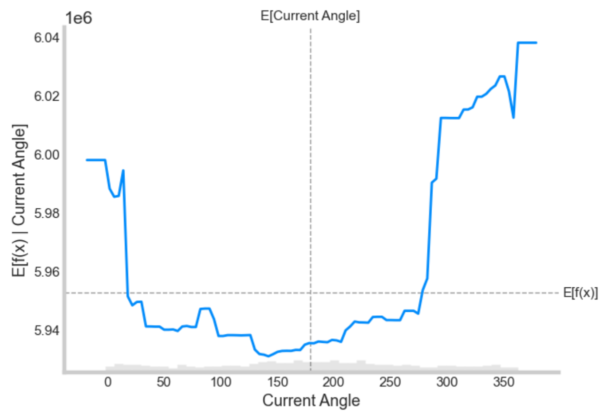

Figure 7. PDP Plot of Current Angle

According to Figure 8, the model's predictions tend to increase when the wind wave is between approximately 200° to 360°. This indicates that when wind waves are within a certain angular range, they create additional resistance likely leading to, which is likely to result in higher fuel consumption and CO2 emissions.

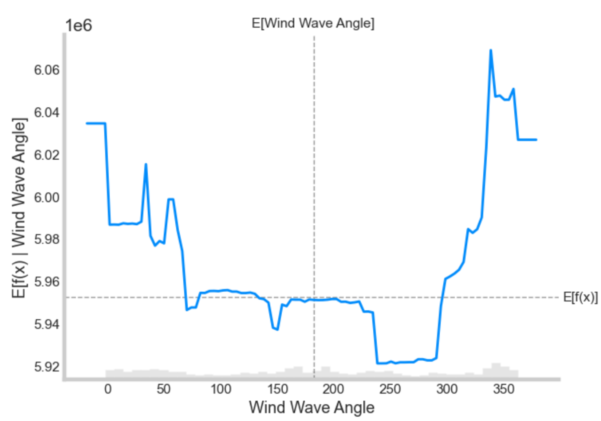

Figure 8. PDP Plot of Wind Wave Angle

According to Figure 9, the model's predictions are highest when the swell wave angle is between approximately 0° and 50°, and they tend to decrease as the angle increases beyond 50°. This suggests that when swell waves are within a specific angular range, they have a significant impact on the ship's progress, increasing resistance, and potentially leading to higher fuel consumption and CO2 emissions.

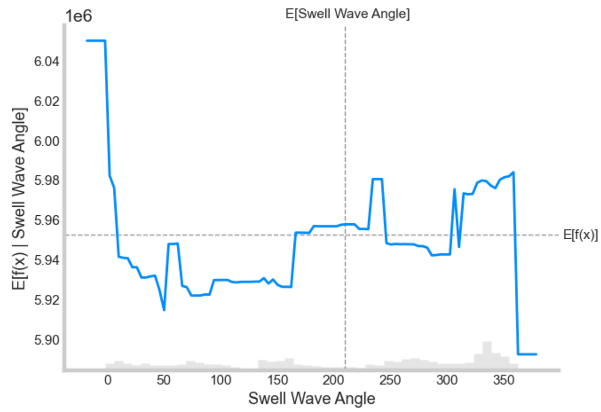

Figure 9. PDP Plot of Swell Wave Angle

4. Conclusion

In this study, a model for predicting ship carbon emissions was developed using XGBoost. The model demonstrated a Mean Absolute Error (MAE) of 89,917 on the test data, which, although higher than that of the training data, remains relatively low given the data's scale. The Mean Absolute Percentage Error (MAPE) was 1.4%, indicating that the model maintains high accuracy on unseen data. The analysis identified speed over ground (SPEED_VG) and relative wind speed (REL_WIND_SPEED) as the most significant variables influencing CO2 emissions, both showing a positive correlation with emissions. This indicates that as the ship's speed over ground and relative wind speed increase, fuel consumption and CO2 emissions tend to rise. Additionally, various environmental factors significantly impact CO2 emissions. A Partial Dependence Plot (PDP) revealed that CO2 emissions increase sharply when the average draft exceeds 22, indicating that deeper submersion increases resistance, leading to higher fuel consumption. Similarly, when leeway exceeds 5°, emissions rise sharply, likely due to increased resistance as the ship maintains its course against environmental forces. CO2 emissions also increase significantly when the current angle exceeds 200°, the wind wave angle is between 200° and 360°, or the swell wave angle is between 0° and 50°. These findings underscore the substantial impact of environmental factors on ship fuel efficiency. Moreover, variables such as pitch (PITCH), roll (ROLL), and rudder angle (RUDDER_ANGLE) exhibited a negative correlation with CO2 emissions. Lower values of these variables were associated with reduced fuel consumption and emissions, suggesting that more stable ship operation enhances efficiency.

In conclusion, this study emphasizes the importance of considering both key operational variables, such as speed over ground and relative wind speed, and a range of environmental factors, including average draft, leeway, current angle, wind wave angle, and swell wave angle, to predict and reduce CO2 emissions from ships. These insights contribute to developing practical strategies for minimizing environmental impact through optimized ship operations.

Future research will aim to incorporate seasonal variations by utilizing year-long data, as a ship's draft and engine performance can be influenced by seasonal factors such as water temperature. This approach will allow for a more accurate representation of how these factors impact ship operations and CO2 emissions.

Acknowledgement

This work was supported Commercialization Promotion Agency for R&D Outcomes(COMPA) and Busan Innovation Institute of Industry, Science & Technology Planning(BISTEP) by grant funded by the

Korea government(MSIT) and a local government(Busan) (No.2023-0487)

References

[1] B. Mary Nathisiya, “Article Title: Adoption of IoT technology in Maritime Sector Adoption of IoT technology in Maritime Sector,” IJATEM.

[2] D. Kaklis, T. J. Varelas, I. Varlamis, P. Eirinakis, G. Giannakopoulos, and C. V Spyropoulos, “From STEAM to Machine: Emissions control in the shipping 4.0 era,” Mar. 07, 2023. DOI: https://doi.org/10.5957/SOME-2023-020.

[3] Y. C. Shih, Y. A. Tzeng, C. W. Cheng, and C. H. Huang, “Speed Optimization in Bulk Carriers: A Weather-Sensitive Approach for Reducing Fuel Consumption,” J. Mar. Sci. Eng., vol. 11, no. 10, 2023, DOI: https://doi.org/10.3390/jmse11102000.

[4] Y. Yuan, X. Wang, L. Tong, R. Yang, and B. Shen, “Research on Multi-Objective Energy Efficiency Optimization Method of Ships Considering Carbon Tax,” J. Mar. Sci. Eng., vol. 11, no. 1, 2023, DOI: https://doi.org/10.3390/jmse11010082.

[5] T. Committee, “80 session of the Marine Environment Protection Committee (MEPC 80),” no. July 2023, pp. 3–7, 2025.

[6] IMO MEPC, “Fourth IMO Greenhouse Gas Study: Executive Summary,” IMO Greenh. Gas Study, vol. 4, no. 1, p. 46, 2020.

[7] IMO, “Resolution MEPC.231(65): 2013 Guidelines for Calculation of Reference Lines for Use with the Energy Efficiency Design Index (EEDI),” Marine Environment Protection Committee, 2013.

[8] IMO, “Resolution MEPC.328(76): Amendments to the Annex of the Protocol of 1997 to Amend the International Convention for the Prevention of Pollution from Ships, 1973, as Modified by the Protocol of 1978 Relating Thereto” (2021 Revised MARPOL Annex VI), Marine Environment Protection Committee, 2021.

[9] J. Barreiro, S. Zaragoza, and V. Diaz-Casas, “Review of ship energy efficiency,” Ocean Eng., vol. 257, no. January, p. 111594, 2022, DOI: https://doi.org/10.1016/j.oceaneng.2022.111594.

[10] K. Alexiou, E. G. Pariotis, and H. C. Leligou, “Sensor Data Quality in Ships: A Time Series Forecasting Approach to Compensate for Missing Data and Drift in Measurements of Speed through Water Sensors,” Designs, vol. 7, no. 2, 2023, DOI: https://doi.org/10.3390/designs7020046

[11] Y. R. Kim, M. Jung, and J. B. Park, “Development of a fuel consumption prediction model based on machine learning using ship in-service data,” J. Mar. Sci. Eng., vol. 9, no. 2, pp. 1–25, Feb. 2021, DOI: https://doi.org/10.3390/jmse9020137.

[12] M. P. Handayani, H. Kim, S. Lee, and J. Lee, “Navigating Energy Efficiency: A Multifaceted Interpretability of Fuel Oil Consumption Prediction in Cargo Container Vessel Considering the Operational and Environmental Factors,” J. Mar. Sci. Eng., vol. 11, no. 11, Nov. 2023, DOI: https://doi.org/10.3390/jmse11112165.

[13] S. Wang, X. Wang, Y. Han, X. Wang, H. Jiang, and Z. Zhang, “Ship Fuel and Carbon Emission Estimation Utilizing Artificial Neural Network and Data Fusion Techniques,” J. Softw. Eng. Appl., vol. 16, no. 03, pp. 51–72, 2023, DOI: https://doi.org/10.4236/jsea.2023.163004.

[14] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., vol. 13-17-August-2016, pp. 785–794, 2016, DOI: https://doi.org/10.1145/2939672.293978.

[15] H. K. M. P. J. Lee, “A Study on the Prediction of Fuel Consumption of Bulk Ship Main Engine Using Explainable Artificial Intelligence,” vol. 47, no. 4, pp. 182–190, 2023.

[16] Y. Ning et al., “Shapley variable importance clouds for interpretable machine learning,” ArXiv, vol. abs/2110.02484, 2021, DOI: https://doi.org/10.48550/arXiv.2110.02484.

[17] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” Adv. Neural Inf. Process. Syst., vol. 2017-December, no. Section 2, pp. 4766–4775, 2017, DOI: https://doi.org/10.48550/arXiv.1705.07874.

[18] A. Guryanov, “Efficient Computation of SHAP Values for Piecewise-Linear Decision Trees,” Proc. ITNT 2021 - 7th IEEE Int. Conf. Inf. Technol. Nanotechnol., pp. 1–4, 2021, DOI: https://doi.org/10.1109/ITNT52450.2021.9649051.

[19] K. Kwon and J. So, “Future Smart Logistics Technology Based on Patent Analysis Using Temporal Network,” Sustain., vol. 15, no. 10, May 2023, DOI: https://doi.org/10.3390/su15108159