

SHAP을 활용한 벌크선 메인엔진 연료 소모량 예측연구

김현주* · 박민규** · † 이지환

*한국조선해양기자재연구원 선임연구원, **부경대학교 석사과정생, † 부경대학교 산업및데이터공학과 교수

A Study on the Prediction of Fuel Consumption of Bulk Ship Main Engine Using Explainable Artificial Intelligence

Hyun-Ju Kim* · Min-Gyu Park** · † Ji-Hwan Lee

*Senior Researcher, Intelligent Convergence Research Team, Korea Marine Equipment Research Institute, Busan 49111, Korea

**Student, Division of Industrial and Data Engineering, Pukyong National University, Busan 48513, Korea

† Professor, Division of Industrial and Data Engineering, Pukyong National University, Busan 48513, Korea

요 약 : 본 연구에서는 벌크 선박의 연료 소비를 예측하기 위해 XGBoost와 SHapley Additive exPlanation (SHAP)을 사용하는 예측 모델을 제안한다. 기존 연구에서도 선박 엔진 데이터와 기상데이터를 활용하였지만 선박 연료소모량 예측 모델에 대한 예측 결과의 신뢰성과 예측 모델 구현에 사용된 변수들에 대한 설명이 부족한 한계가 있었다. 이러한 문제를 해결하기 위해 본 연구에서는 XGBoost와 SHAP를 사용하여 예측 모델을 개발하였다. 이 연구는 연구 배경, 범위, 관련 규정 및 이전 연구들, 그리고 연구 방법론에 대한 소개를 제공하며, 또한 벌크 선 데이터 정제 방법과 예측 모델 결과의 검증을 설명한다.

핵심용어 : 벌크선, 연료 소비 예측, XGBoost, SHAP, 외부 기상 데이터

Abstract : This study proposes a predictive model using XGBoost and SHapley Additive exPlanation (SHAP) to estimate fuel consumption in bulk carriers. Previous studies have also utilized ship engine data and weather data. However, they lacked reliability in predicted results and explanations of variables used in the fuel consumption prediction model implementation. To address these limitations, this study developed a predictive model using XGBoost and SHAP. It provides research background, scope, relevant regulations, previous studies, and research methodology. Additionally, it explains the data cleaning method for bulk carriers and verifies results of the predictive model.

Key words : bulk carriers, fuel consumption prediction, XGBoost, SHAP, external weather data.

1. 서 론

최근 전 세계적인 탄소중립 정책에 따라 국제해사기구(International Maritime Organization; IMO)는 2050년까지 온실가스 감축을 위한 전략을 수립하고 있다. 특히 최근에 개최된 해양환경보호위원회(Maritime Environment Protection Committee; MEPC) 제 80차 회의에서 2050년 해운 분야 탄소 배출 감축 목표를 기존 2008년 대비 50%에서 100%로 상향하는 'Net-zero' 초기전략 개정이 가결됐다(IMO, 2023)

따라서 탄소 배출 감축을 2008년 대비 100% 감축하기 위해 국제해사기구에서는 다양한 규제를 시행하고 있다.

대표적인 규제로 에너지효율설계지수(Energy Efficiency Design Index; EEDI), 에너지효율운항지수(Energy Efficiency Operational Index; EEOI), 탄소집약도지수(Carbon Intensity Indicator; CII)와 같은 환경 규제를 도입하여 선박의 배기가스

배출량을 줄이기 위한 노력을 하고 있다.

탄소집약도지수(Carbon Intensity Indicator; CII)는 선박의 운송 작업량 당 CO₂ 배출량 평균을 나타는 지표로 2023년부터 기존의 국제 운항선을 대상으로 연간 탄소배출집약도 보고를 의무화하였으며 규제를 미충족 선박에는 운항을 중단하도록 강력히 규제하고 있다. 특히, 탄소집약도지수(Carbon Intensity Indicator; CII)를 충족하기 위해서는 연간 연료유 소모량을 줄이는 것이 가장 중요하다. 따라서, 정확한 연료소모량 예측을 통한 효율적인 선박의 관리는 환경규제 충족을 위한 핵심 요소라고 할 수 있다.

이러한 선박의 연료소모량 예측을 위하여 수집된 선박의 빅데이터를 활용한 다양한 연구들이 수행되었다. 기존의 많은 연구에서는 예측이 어려운 선박의 외부 기상 환경 데이터를 제외한 후 선박의 운항 데이터만 활용한 예측 모델을 구현하

† Corresponding author : 정희원, jihwan@pknu.ac.kr 051)629-6492

* 정희원, hjkim94@komeri.re.kr 051)400-5285

** 정희원, mingyu7539@gmail.com 051)629-6492

었다. Kim et al.(2019)은 주성분 분석(Principal Component Analysis; PCA)을 선박의 엔진 데이터와 외부 기상 환경 데이터를 활용하여 회귀 모형에 적용하여 선박의 연료 소모량을 예측 모델을 구현하였다. Uyanik et al.(2020)은 선박의 엔진 데이터만을 활용하여 pearson 상관관계 분석과 회귀 모델을 사용하여 연료 소모량을 예측하였다. 이러한 경우 모형의 과적합 문제가 많이 발생하였기 때문에 최근에는 선박의 운항 데이터와 외부 기상 환경 데이터를 동시에 활용하거나 외부의 기상 환경 데이터만을 활용하여 연료소모량 예측 모델을 구현하고 있다. Shin et al.(2021)은 선속과 해양기상정보가 선박 연료소모량에 미치는 영향을 분석하고자 상관관계 분석을 통해 실선 실험 해석을 진행하였다. 하지만 상관관계 분석은 변수들 간의 선형적인 관계를 파악할 수 있으나 해당 변수가 예측 결과에 기여하는 정도는 파악할 수 없다는 단점이 존재한다.

정리하면 기존의 연구에서는 수집된 선박 운항 데이터와 외부 기상 환경 데이터를 머신러닝, 딥러닝 기법을 활용해 예측 모형의 정확도를 향상시켰으나 예측 모형 내부에서 일어나는 의사 결정과 예측 모형 구현에 활용된 변수와 예측 모형의 결과에 대한 영향력을 설명할 수 없었다. 즉, 정확한 연료소모량 예측 모형을 만들었지만 변수로 사용된 선박의 속도, 홀수, 풍향, 풍속 등 개별 요소별로 예측 결과에 긍정적 혹은 부정적으로 영향을 미치는지는 확인할 수 없는 문제가 있었다.

따라서 본 연구에서는 이러한 문제를 해결하기 위해서 정확한 모형을 예측하는 것을 넘어서 예측 모형에 대하여 각각의 변수가 가지는 중요도를 분석하고자 한다. 먼저 실제 운항 중인 벌크선박의 데이터를 수집하여 XGboost(eXtreme Gradient Boosting)를 사용하여 선박의 메인엔진 연료소모량 예측 모형을 구현한다. 다음으로 예측 모형에 사용된 변수들이 모형에 어떻게 작용하였는지 SHapley Additive exPlanation(SHAP)을 통한 결과를 분석한다. 마지막으로 전체 운항 구간과 연료소모량이 가장 컸던 운항 구간에 대한 SHapley Additive exPlanation(SHAP)결과를 비교 분석하여 연료소모량에 미치는 변수들의 영향도를 파악하고자 한다.

2. 연구 방법

2.1 연구 절차

본 연구에서는 Fig. 1 과 같은 순서로 연구를 진행한다. 수집된 선박의 Raw Data 정제를 포함하여 유의미한 결과 도출을 위하여 데이터 표준화, 변수들의 조합을 통한 새로운 변수 생성과 같은 전처리 작업을 진행하였다. 다음으로 예측 모형 생성을 위하여 데이터를 학습과 테스트를 위하여 8대 2의 비율로 분리하였으며 XGboost(eXtreme Gradient Boosting)모형을 적용하여 선박 연료소모량 예측 모형을 개발하였다. 다음으로 예측 모형 평가를 통하여 모형의 정확도를 확인하였다. 마지막으로 SHAP values를 이용하여 개발한 예측 모형에

사용된 변수들이 모형의 예측에 어떤 영향을 미치는지를 설명하였다.

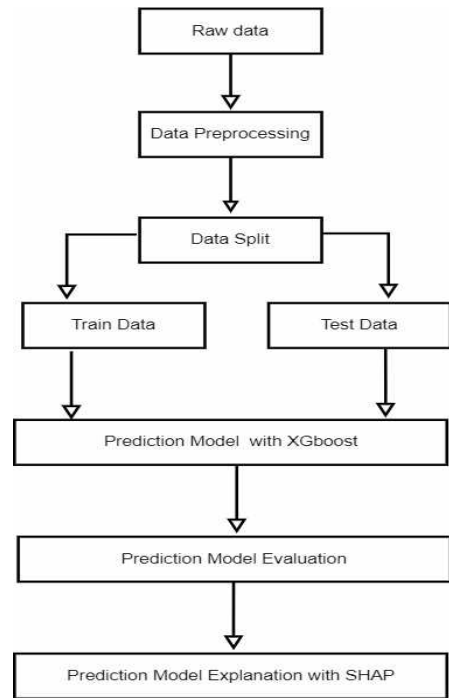


Fig. 1 Flow chart of the study

2.2 분석 대상 및 데이터 설명

2.2.1 분석 대상

본 연구에서는 사용된 대상 선박은 16만 톤급의 대형 화물 선박이다. 대형 화물 선박의 경우에는 연료비가 전체 운항 경비의 약 50~60%를 차지하기에 많은 해운회사들이 에너지 소비를 줄이기 위해 노력하는 선종 중 하나이다(Stopford, 2009). 최근 국제적인 선박의 환경규제에서 5,000톤급 이상의 국제 원양 선박에 대해서는 연료소모량이 중요 지표로 작용하는 규제 등급을 산정하는 점을 고려하여 해당 선박으로 선정하였다. 또한, 화물 선박은 주로 대서양과 태평양을 연결하는 긴 항로를 항해하는 경우가 많아 데이터 분석시에 안정적인 데이터양을 수집할 수 있다는 것을 고려하여 최종 분석 대상 선박을 Table 1과 같이 선정하였다.

Table 1 Vessel Specification

Specification	Measurement
Type of ship	(BULK) CARRIER
L.O.A (m)	333
Design Speed (knots)	14
Max Draft (m)	22
Gross Tonnage (ton)	160,000

2.2.2 데이터 설명

본 연구에서는 대상 선박이 2020년 01월 01일부터 2020년 06월 30일까지의 6개월 간 선박 운항 데이터를 연구에 사용하였다. 선박 운항 데이터는 10초 주기로 수집된 데이터를 10분 주기로 압축한 데이터를 활용하여 연구를 진행하였다. 수집된 선박 운항 데이터 변수의 총개수는 190개이나 본 연구에서는 19개의 변수를 사용하였으며 Table 2와 같이 나타내었다.

Table 2 Original Variable

	Name	Description	Unit
1	MEI FOC	no.1 Main Engine Fuel Oil Consumption	kg/h
2	Speed Over Ground	Speed over ground	knots
3	Fore Draft	Vertical distance between the waterline and the bottom of the hull measured at the perpendicular of the bow	m
4	Aft Draft	Vertical distance between the waterline and the bottom of the hull measured at the perpendicular of the stern	m
5	Port Draft	Vertical distance between the waterline and the bottom of the hull measured at the perpendicular of the port side (left)	m
6	Starboard Draft	Vertical distance between the waterline and the bottom of the hull measured at the perpendicular of the starboard side(right)	m
7	Ship Heading	Compass direction in which the ship's bow is pointed	degree
8	Course Over Ground	Direction of which the vessel is moving	degree
9	Rudder Angle	angle of rudder	degree
10	Wind Wave height	Significant height of wind waves	m
11	Wind Wave Direction	Direction of wind waves	degree
12	Wind Wave Period	Mean period of wind waves	s
13	Current Speed	Directional movement of seawater	m/s
14	Current Direction	Direction of seawater	degree
15	Swell Wave Height	Significant height of swell waves	m
16	Swell Wave Direction	Direction of swell waves	degree
17	Swell Wave Period	Mean period of swell waves	s
18	Relative Wind Speed	Speed of the apparent wind	m/s

19	Relative Wind Direction	Direction of the apparent wind	degree
----	-------------------------	--------------------------------	--------

2.3 데이터 정제 및 전처리

선박에서는 센서 노후화, 기계적 이상, 통신 문제 등 다양한 이유로 수집된 센서 데이터 오류가 발생한다. 본 연구에 활용된 운항 데이터는 오류나 결측치가 발생하면 -9999와 error와 같이 기록되며 본 연구에서는 이러한 오류 데이터를 삭제하는 정제작업을 수행하였다.

또한 Table 2의 변수간의 일부 조합을 통하여 새로운 변수를 생성하였다. 변수간의 조합을 통해 변수간의 상호작용을 파악할 수 있으며 이는 결국 예측 모델의 정확도를 높이고 예측 모델에 대한 설명력을 높이는 장점이 있다. 본 연구에서는 선박의 연료 소모량 예측에 사용한 변수가 연료 소모량에 어떤 영향을 미치는지를 설명하고자 하였다.

먼저 GPS(Global Positioning System)로부터 측정되는 선박의 실제 침로인 COG(Course Over Ground)와 Gyroscope로부터 측정되는 Ship Heading의 산술차이를 Leeway라고 식(1)과 같이 추가하였다(Yoon, 2005). 또한 본 연구에서는 선박이 해상에서 발생하는 조류 영향으로 선박의 이동방향과 항로가 일치하지않는 정도를 Tideway라고 하여 조류의 방향(Current Direction)과 Ship heading 산술차이로 식(2)와 같이 정의하였다. Leeway와 Tideway를 Fig. 2와 같이 나타내었다.

외부 환경 조건에 의한 선박 연료소모량에 미치는 영향도를 세부적으로 설명하고자 풍랑, 너울, 조류 등의 방향과 Ship heading과의 산술차이를 식(3)~(5)과 같이 새로운 변수로 추가하였다. 대표적으로 Swell Wave Angle을 Fig. 3와 같이 나타냈으며 나머지 외부 환경 조건에 의한 변수들도 동일한 방식을 사용하였다.

선수, 선미, 좌현, 우현의 흘수의 산술평균으로 계산한 평균 흘수를 식(6)과 같이 새로운 변수로 추가하였으며, 트림(Trim)과 힐(Heel)을 절대값 처리하여 식(7)~(8)과 같이 추가하였다.

$$Leeway = Course\ Over\ Ground - Ship\ Heading \quad (1)$$

$$Tideway = Current\ Direction - Ship\ Heading \quad (2)$$

$$Wind\ Wave\ Angle = Wind\ Wave\ Direction - Ship\ Heading \quad (3)$$

$$Swell\ Wave\ Angle = Swell\ Wave\ Direction - Ship\ Heading \quad (4)$$

$$Relative\ Wind\ Angle = Relative\ Wind\ Direction - Ship\ Heading \quad (5)$$

$$AvgDraft = \frac{Fore\ Draft + Aft\ Draft + Port\ Draft + Starboard\ Draft}{4} \quad (6)$$

$$Trim = |Fore\ Draft - Aft\ Draft| \quad (7)$$

$$Heel = |Port\ Draft - Starboard\ Draft| \quad (8)$$

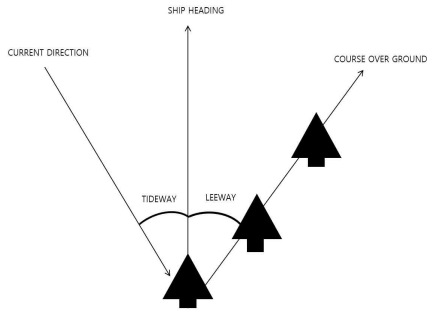


Fig. 2 Leeway and Tideway

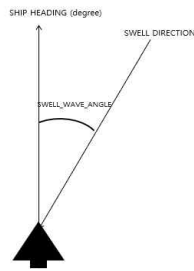


Fig. 3 Swell Wave Angle

선박의 연료 소모량 예측에 선박이 항해할 때 외부 환경 조건에 의한 영향도를 설명하고자 하였기에 변수 데이터 중에 Ship Heading과 같이 360도로 측정되는 원형 구조의 변수는 모두 0° ~ 90° 사이의 예각으로 Table 3 와 같이 처리하였다. 이는 원형 특성을 가지는 각(degree) 데이터 특성을 고려하여 외부의 환경 조건인 풍랑, 너울, 조류 등의 방향이 Ship Heading에 미치는 영향을 보다 더 정확하게 파악하고자 위함이다.

Table 3 conversion of angles

	Range	Processing	Unit
1	90° <angle < 180°	angle - 90°	degree
2	180° <angle < 270°	angle - 180°	degree
3	270° <angle < 360°	angle - 270°	degree

따라서 기존의 19개의 변수에서 데이터 전처리를 통해 학습에 사용된 최종 변수는 총 17개로 Table 4와 같이 정리하였다.

Table 4 final selected variables

	Name	Description	Unit
1	ME1 FOC	-	kg/h
2	Speed Over Ground	-	knots
3	AvgDraft	Arithmetic mean of fore, aft, port, starboard draft	m
4	Trim	The longitudinal slope of a ship	m
5	Heel	The horizontal slope of a ship	m
6	Leeway	Angular difference between Course of Ground and Ship	degree

		heading	
7	Tideway	Angular difference between Current Direction and Ship heading	degree
8	Wind Wave Angle	Angular difference between Wind Wave Direction and Ship heading	degree
9	Swell Wave Angle	Angular difference between Swell Wave Direction and Ship heading	degree
10	Relative Wind Angle	Angular difference between Relative Wind Direction and Ship heading	degree
11	Rudder Angle	-	degree
12	Wind Wave Height	-	m
13	Wind Wave Period	-	s
14	Current Speed	-	m/s
15	Swell Wave Height	-	m
16	Swell Wave Period	-	s
17	Relative Wind Speed	-	m/s

2.4 데이터 분석 조건

본 연구에서는 활용되는 선박의 MCR(Maximum Continuous Rating)의 75~83% 구간의 메인엔진의 RPM(revolutions per minute)과 Speed를 고려하여 메인엔진 RPM(revolutions per minute) 50 r/min 이상, 대지속도(Speed Over Ground) 10 knot 이상의 선박이 안정적으로 항해하는 구간의 데이터를 활용하였다. 이러한 구간에서는 선박이 경제적이고 안정적이라는 이유로 일반적으로 일정 방향을 유지하며 항해하는 구간이라고 볼 수 있다. 목적지에 도달하기 위하여 방향을 바꾸는 경우를 제외한다면 Leeway가 크게 발생했다면 항해 구간에서 발생하는 외력의 영향이라고 추측해 볼 수 있기 때문이다.

3. 분석 결과

3.1 XGboost 적용 선박 연료소모량 예측 모델

3.1.1 XGboost(eXtreme Gradient Boosting) 이론

부스팅(Boosting)은 머신러닝에서 널리 사용되는 앙상블 기법 중 하나이다(Optiz, 1999). 대표적인 부스팅 알고리즘 기법인 그래디언트 부스팅(Gradient Boosting)과 XGboost(eXtreme Gradient Boosting)이 있다. 그래디언트 부스팅(Gradient Boosting)은 약한 학습기(weak learner)인 결정 트리를 연속적으로 학습하여 강력한 앙상블 모델을 만드는 알고리즘이다.

XGboost(eXtreme Gradient Boosting)모델은 Gradient Boosting 알고리즘을 기반으로 약한 학습기(weak learner)인

의사 결정 트리 여러 개를 조합하여 단일 의사결정 트리보다 강력한 앙상블 모델을 만들도록 확장하고 개선시킨 모델이다 (Chen, 2016). 각 트리는 서로 다른 부분 데이터셋으로 이전 모델의 오류를 보완하도록 순차적으로 학습되고 최종 예측은 이들의 예측 결과를 결합하여 도출되는 방식이다. 따라서 이러한 방법은 단일 모델보다 더 좋은 예측 성능을 제공할 수 있다는 장점이 있다. 또한 순차적으로 모델을 학습하는 점 때문에 모델의 과적합을 보정하며 빠른 학습 속도도 장점이다. 이러한 점으로 인해 다양한 머신러닝 문제에서 많이 활용되고 있으며 본 연구에서도 최종 모델로 채택하였다.

3.1.2 예측 모델 정확도

본 연구에서는 XGBoost 기반으로 연료소모량 예측 모델을 생성하였으며, GridSearchCV(Grid Search Cross-Validation)를 활용하여 하이퍼파라미터 튜닝을 진행하였다. 모델의 성능 평가 결과는 과적합(overfitting) 여부를 확인하기 위해 예측 모델의 성능을 학습데이터와 테스트데이터로 각각 모델을 평가하였으며 회귀 모델 성능 평가에 대표적으로 사용되는 평균제곱근오차 (Root Mean Squared Error, RMSE), 평균절대오차(Mean absolute error, MAE), R^2 (R-squared), 평균백분율오차 (Mean Absolute Percentage Error, MAPE)를 사용하여 결과를 Table 5와 Table 6과 같이 정리하였다. 학습용 데이터와 평가용 데이터를 사용해 모델을 각각 평가한 결과 모델의 과적합 문제가 발생하지 않았음을 알 수 있었다.

통계학 측면에서 R^2 값이 0.7 이상인 경우 좋은 모델로 평가받는다(Moor and Kirkland, 2007). 따라서 본 연구에서 평가용 데이터를 사용해 모델의 성능을 평가한 결과 모델의 정확도를 나타내는 R^2 값이 0.9512를 가지며, 오차를 나타내는 MAPE는 0.0158를 가지는 것을 확인하여 연료소모량 예측 모델이 우수한 예측 정확도를 가짐을 확인할 수 있었다. Fig. 4와 Fig. 5에서 보면 모델의 예측 성능을 산점도를 확인할 수 있다. 붉은색 선이 예측 모델을 나타내며 파란색으로 분포된 점은 데이터를 나타낸다. Fig. 4는 학습용 데이터에 대한 예측 모델과의 관계를 볼 수 있으며, Fig. 5는 평가용 데이터에 대한 예측 모델과의 관계를 확인할 수 있으며 예측 모델이 우수한 성능을 가짐을 시각적으로 알 수 있다.

Table 5 The performance evaluation of the model using the training dataset

	RMSE	MAE	R^2	MAPE
1	9.2580	6.3690	0.9811	0.0033

Table 6 The performance evaluation of the model using the test dataset

	RMSE	MAE	R^2	MAPE
1	62.4345	31.3686	0.9512	0.0158

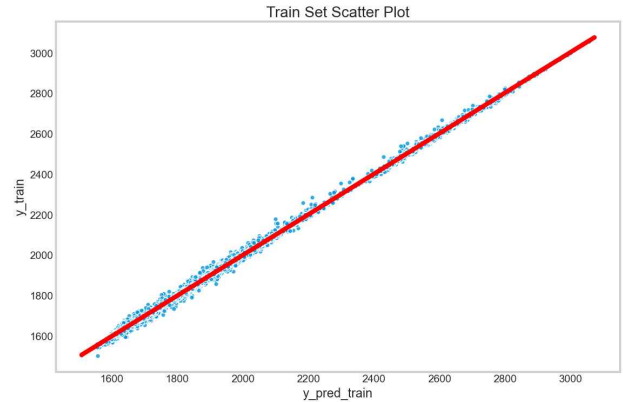


Fig. 4 The Scatter Plot of the model evaluated on the training dataset

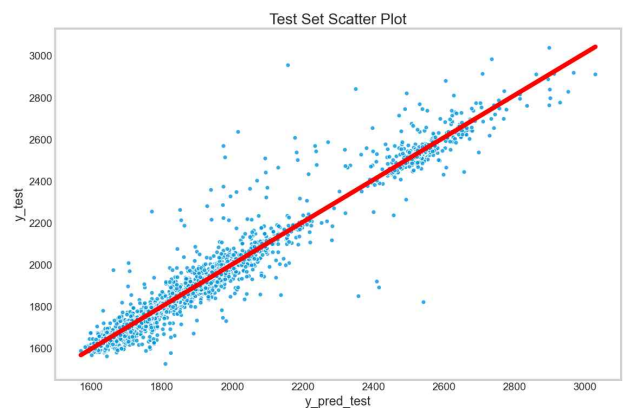


Fig. 5 The Scatter Plot of the model evaluated on the test dataset

3.2 SHapley Additive exPlanation(SHAP)을 통한 모델 해석

3.2.1 XAI(Explainable Artificial Intelligence) 이론

인공지능은 자동화 및 높은 정확도를 통해 기존의 탐지기법에 비해 많은 관심을 받고 있지만, 블랙박스(Black-box) 특성으로 인해 현장에서는 적용하기 어려움을 가지고 있다 (Tjoa, 2020). XAI(Explainable Artificial Intelligence)는 이러한 인공지능 모델의 동작 방식을 설명 가능하도록 만드는 기술이다(Lipton, 2016). XAI는 인간이 이해할 수 있는 방식으로 작동 원리와 의사 결정 과정을 설명하는 AI 시스템을 개발하는 것을 목적으로 한다. AI는 머신러닝 모델의 예측 결과를 해석하고, 모델의 내부 동작을 분석하여 인과적인 관계나 패턴 등을 추론할 수 있다. XAI의 이론은 크게 모델 내부 설명력과 모델 외부 설명력으로 구분할 수 있으며, 모델 내부 설명력은 모델 자체가 설명가능한 모델인지 여부를 평가하며, 모델 외부 설명력은 모델을 해석하는 기술과 방법을 의미한다.

모델 외부 설명력은 다양한 시각화 기술, 질문-답변 형태의 인터페이스, 자연어 처리 등의 기술로 모델을 해석하는 것을 의미하여 데이터의 특징이나 모델의 동작 방식 등을 이해하

기 쉽도록 가시화하여 제공한다. 질문-답변 형태의 인터페이스는 사용자가 모델을 직접 조작하면서 모델의 결과를 해석할 수 있도록 하며 자연어 처리는 모델의 결과를 인간이 이해하기 쉬운 자연어 형태로 변환하여 제공한다. 즉, XAI(Explainable Artificial Intelligence) 기법은 인간과 인공지능 간의 의사소통을 지원하며 모델의 결과가 설득력있게 전달하고, 모델의 내부 동작이 명확하게 이해시킬 수 있도록 활용되는 방식이다.

3.2.2 SHAP (SHapley Additive exPlanations) 이론

SHAP (SHapley Additive exPlanations)은 XAI (Explainable Artificial Intelligence) 분야에서 사용되는 모델 설명 기법이다(Lundberg, 2017). SHAP value는 합리적인 분배 방식을 제공하는 게임 이론에서 유래한 개념으로 변수들의 중요도를 평가할 수 있다. SHAP은 각 변수가 예측 결과에 얼마나 기여하는지를 계산하여 음의 관계, 양의 관계까지 영향력의 방향을 수치적으로 설명할 수 있다(Lloyd, 1953).

즉, SHAP을 사용하여 모델을 분석하면 각각의 입력변수의 값이 변할 때 예측값이 어떻게 변화하는지를 해석할 수 있다. SHAP은 변수들 간의 상호작용을 고려하여 예측 결과 설명을 제공한다. 이를 통해, 모델의 개별 예측 결과를 설명할 뿐 아니라, 전체 데이터에 대한 모델의 해석 가능성을 높일 수 있다는 장점이 있다.

3.2.3 Global Importance를 통한 모델 해석 결과

Global Importance는 SHAP value의 평균화를 통해 계산된 값이다(Mengnan, 2019). 이를 통해 모델의 예측에 어떤 변수들이 평균적으로 중요하게 작용하는지를 확인할 수 있다. Fig. 6 같이 연료소모량 예측 모델에 영향을 미친 변수들의 SHAP value 평균을 확인할 수 있으며 선박의 대지속도(Speed Over Ground), 선박의 상대 풍속(Relative wind speed), 선박의 평균흘수(Average Draft), 트림(Trim), 힐(Heel), 너울의 높이(Swell Wave Height) 순서로 비교적 높은 영향도를 확인하였다.

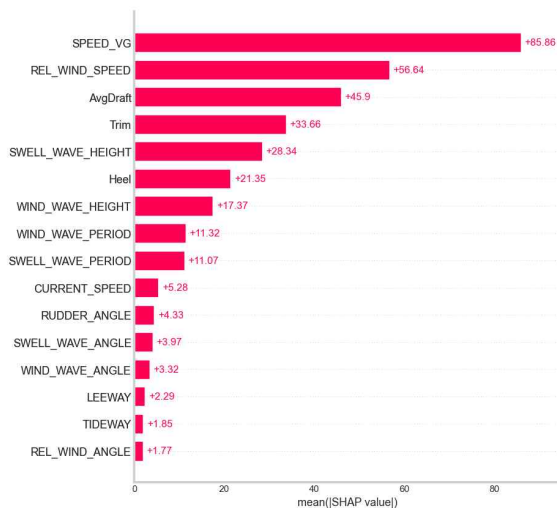


Fig. 6 Global Importance

3.2.4 SHAP Value를 통한 모델 해석 결과

SHAP value는 개별적인 변수들이 모델 예측에 기여하는 정도를 나타내는 값으로 Fig 7에 제시된 SHAP value plot은 각 변수의 SHAP Value와 해당 값이 모델의 예측 결과에 미치는 영향을 색으로 구분하여 보여준다. x축의 0을 기준으로 좌측은 변수들의 값이 감소하는 것을 나타내며 우측은 변수들의 값이 증가하는 것을 나타낸다. 또한 붉은색은 변수들은 모델의 예측 결과를 증가시키는 긍정적 영향을 미치며, 파란색은 변수들은 모델의 예측 결과를 감소시키는 부정적 영향을 미친다.

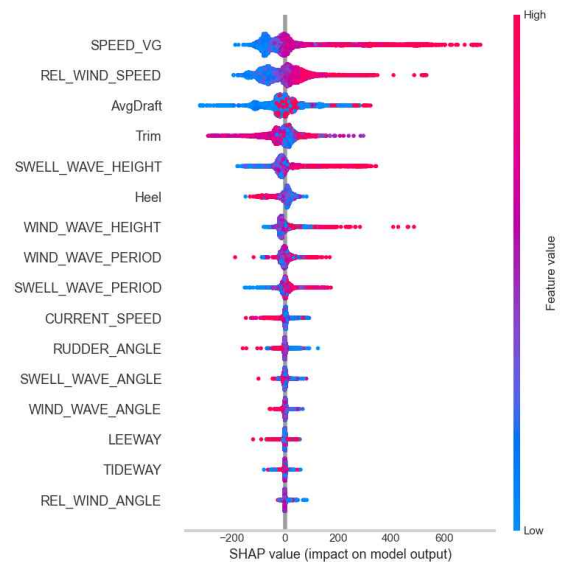


Fig. 7 SHAP values plot

Fig. 7을 통해 종합적으로 정리하여 분석하면 선박의 대지속도(Speed Over Ground)가 클수록, 상대 풍속(Relative Wind Speed)이 클수록, 너울의 높이(Swell Wave Height)가 높을수록, 선박 연료소모량이 증가할 가능성이 있다는 것을 확인할 수 있다.

3.2.5 PDP(Partial dependence plot)를 통한 변수의 의존성 분석

Dependence plot은 변수와 모델의 예측값 간의 관계를 시각적으로 보여주는 기법이다. 변수의 값이 변할 때 모델의 예측값이 어떻게 변화하는지를 확인할 수 있다. 마찬가지로 Partial dependence plot (PDP)은 변수와 모델의 예측값 간의 관계를 시각화하는 기법이나 Dependence plot과 다르게 하나의 변수가 다른 변수와 상호작용하지 않는 가정하에 해당 변수 범위를 변화시켰을 때 모델 예측값이 어떻게 변화하는지 볼 수 있다(Greenwell, 2017). 이러한 기법을 함께 분석하여 더욱 정확한 모델 해석을 할 수 있다.

앞서 Fig. 6을 통해 선박의 연료소모량 예측에 평균적으로 중요하게 영향을 미친 변수 3가지인 선박의 대지속도(Speed Over Ground), 선박의 상대 풍속(Relative wind speed), 선박

의 평균흘수(Average Draft)에 대하여 Dependence plot과 PDP(Partial dependence plot) 분석을 진행하였다.

Fig. 8에서 보면 대지속도(Speed Over Ground)가 붉은색으로 두드러지게 나타나는 15knot 이상일 경우 선박 연료소모량 증가에 크게 기여함을 알 수 있다. Fig. 9에서 보면 대지속도(Speed Over Ground) 13knot 이상부터 선박 연료소모량 증가에 기여함을 알 수 있다.

Fig. 10에서 보면 상대 풍속(Relative Wind Speed)가 20knot 이상일 경우 선박 연료소모량 증가에 크게 기여하고 Fig. 11에서 보면 11knot 이상부터 선박 연료소모량 증가에 기여함을 알 수 있다.

Fig. 12에서 보면 평균흘수(Average Draft)가 22m 이상일 경우 선박 연료소모량 증가에 크게 기여하고 Fig. 13에서 보면 17m 이상부터 선박 연료소모량 증가에 기여함을 알 수 있다.

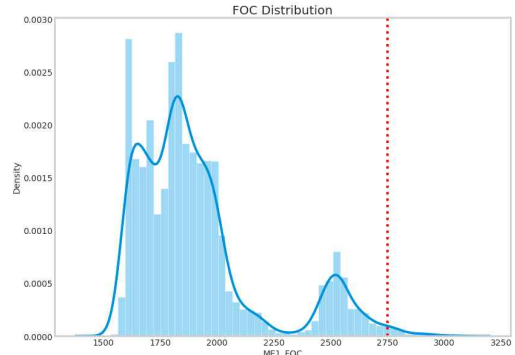


Fig. 14 ME1_FOC Distribution

ME1_FOC가 2750 이상인 구간의 데이터 수는 152개로 나타났다으며 Fig 15의 붉은색 점들과 같이 남아프리카의 케이프타운을 근처로 밀집한 것으로 확인하였다.

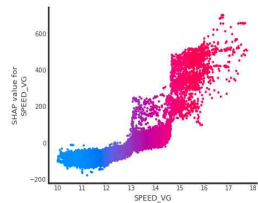


Fig. 8 Dependence plot of speed over ground

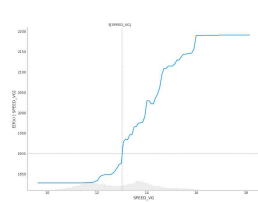


Fig. 9 PDP plot of speed over ground

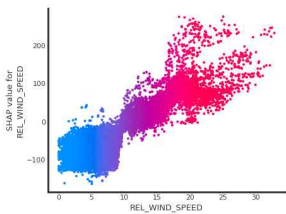


Fig. 10 Dependence plot of Relative wind speed

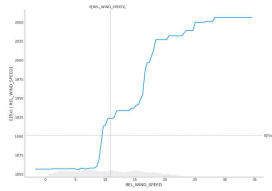


Fig. 11 PDP plot of Relative wind speed

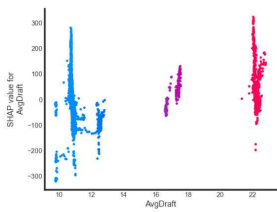


Fig. 12 Dependence plot of AvgDraft

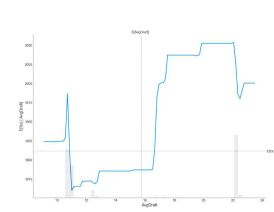


Fig. 13 PDP plot of AvgDraft



Fig. 15 Map showing observations with high ME1_FOC (red dots)

SHAP value의 평균을 계산하여 남아프리카 케이프타운 지역과 전체 지역의 SHAP value를 비교하기 위하여 이를 데이터프레임 형태로 만들어 Table 7과 같이 SHAP 평균값을 비교하였다.

전체 지역에서는 선박의 대지속도(Speed Over Ground)가 14.258555로 가장 높은 SHAP value를 가지고 트림(Trim)이 8.568977로 다음으로 높은 SHAP value를 가지는 것을 확인하였다.

하지만 연료소모량이 크게 발생한 남아프리카 케이프타운 지역에서는 선박의 상대 풍속(Relative wind speed), 선박의 대지속도(Speed Over Ground), 풍랑의 주기(Wind Wave Period), 너울의 주기(Swell Wave Period), 조류의 속도(Current Speed)의 SHAP Value가 특히 크게 나타났으며, 이를 통해 해당 변수들이 연료소모량을 증가시키는 요소들로 발견됨을 알 수 있다. 이러한 결과를 Fig 16의 South africa(과란색) 그래프를 통해서도 확인할 수 있다.

3.2.6 연료소모량이 높은 특정 관측치와의 비교분석

본 연구에서는 연료소모량이 크게 발생한 관측데이터와 전체 데이터간의 SHAP Value 비교를 진행하였다. Fig 14와 같이 ME1_FOC(no1. Main Engine Fuel Oil Consumption) 분포를 통해 ME1_FOC가 2750 이상인 높은 구간의 데이터를 분석하여 ME1_FOC가 높게 나타나는 원인을 분석하였다.

Table 7 Comparison of shap value

	Name	South africa	All
1	Relative Wind Speed	236.029953	-10.235121
2	Speed Over Ground	190.759094	14.258555

3	Wind Wave Height	121.321915	0.232600
4	Wind Wave Period	96.488419	1.981623
5	Swell Wave Period	70.867287	1.318987
6	Current Speed	27.607906	0.989486
7	Trim	24.453243	8.568977
8	Rudder Angle	17.998415	0.696779
9	Heel	11.916217	-2.658451
10	Leeway	11.823912	-0.035883
11	Wind Wave Angle	7.019795	-0.264497
12	Swell Wave Angle	1.337704	-0.662614
13	Tideway	0.234860	-0.023434
14	Relative Wind Angle	-0.091090	-0.134646
15	Swell Wave Height	-9.667458	0.607939
16	AvgDraft	-39.065880	-1

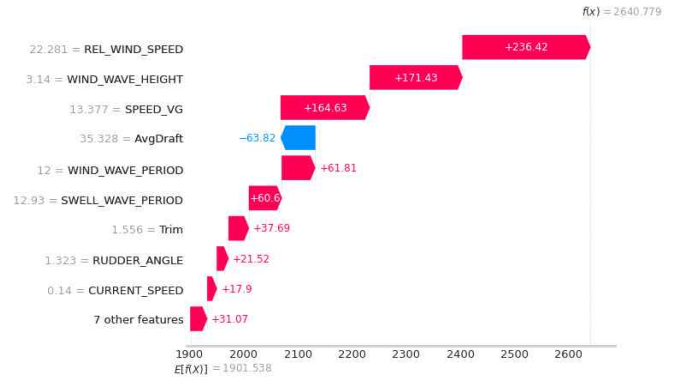


Fig. 17 SHAP Waterfall Plot

4. 결론 및 향후 연구

본 연구에서는 16만톤급의 벌크선 데이터를 사용하여 선박 메인엔진의 연료소모량 예측 모델을 구현하였고 SHAP(SHapley Additive exPlanations)을 통하여 예측 모델의 결과에 대해 변수의 작용을 해석하였다. 모델의 예측 정확도는 R^2 (R-squared) 값은 0.9512, 평균백분율오차 (Mean Absolute Percentage Error, MAPE) 값은 0.0158로 우수한 모델 정확도를 확인할 수 있었다.

본 연구에서는 SHAP value 평균을 통하여 연료소모량 예측 모델에 중요한 영향을 미치는 변수를 선박의 대지속도 (Speed Over Ground), 선박의 상대 풍속(Relative wind speed), 선박의 평균흘수(Average Draft), 트림(Trim), 횡(Heel), 너울의 높이(Swell Wave Height)와 같은 순서로 확인하였다.

변수간의 상호의존성을 분석한 결과 대지속도(Speed Over Ground)가 15knot 이상, 상대 풍속(Relative Wind Speed)이 20knot 이상, 평균흘수(Average Draft)가 22m 이상일때 연료소모량 예측값 증가에 크게 기여함을 알 수 있었으며, 다른 변수들과의 관계를 배제하고 개별 변수들만의 의존성을 분석한 결과 대지속도(Speed Over Ground)가 13knot 이상, 상대 풍속(Relative Wind Speed)이 11knot 이상, 평균흘수(Average Draft)가 17m 이상일 때 선박 연료소모량 증가에 기여함을 알 수 있다.

또한, 본 연구에서는 선박 연료소모량이 가장 크게 나타난 남아프리카 지역과 전체 운항지역의 변수의 SHAP value 차이를 비교하였다. 남아프리카 지역에서는 연료소모량에 특히 외부 환경 조건에 대한 변수들이 더 큰 영향을 미쳤다는 것을 확인할 수 있었다. 특히 ME1_FOC가 가장 높았던 포인트에서는 선박의 상대 풍속(Relative wind speed), 풍랑의 높이 (Wind Wave Height), 선박의 대지속도(Speed Over Ground)가 연료소모량을 높이는데 기여했다는 점을 분석 결과를 통해 확인하였다.

이러한 결과는 선박 운항 관리에 있어서 매우 유용하며, 선

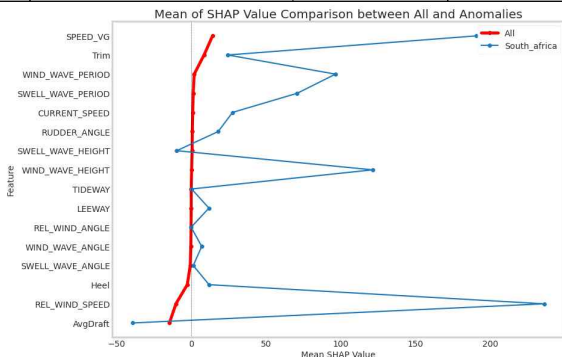


Fig. 16 Mean of SHAP Value comparison between All and South africa

또한 ME1_FOC가 2750 이상인 구간 중에서도 가장 높은 개별 포인트의 SHAP Waterfall Plot을 통해 변수들의 기여도를 추가로 분석하였다. Fig 17에서 보면 선박의 상대 풍속 (Relative wind speed), 풍랑의 높이(Wind Wave Height), 선박의 대지속도(Speed Over Ground), 풍랑의 주기(Wind Wave Period), 너울의 주기(Swell Wave Period), 트림(Trim), 타각(Rudder Angle), 조류의 속도(Current Speed)가 ME1_FOC를 높이는데 기여함을 알 수 있다.

박 운항을 보다 효율적으로 관리하여 연료비를 절감하고, 친환경적인 선박 운항을 실현할 수 있는 데에 기여할 수 있으며 이번 분석에서 사용한 모델과 해석 방법론은 다음과 같은 유사한 문제에 적용될 수 있을 것으로 기대된다. 높은 정확도를 가진 연료소모량 예측 모델을 수립하고 해당 모델에 영향을 미친 변수들의 결과를 분석하여 수치적으로 제시한다는 점에서 해운사에서 선박의 속도 관리, 항로 관리, 항해 계획 수립, 선박 에너지 소모량 모니터링 등 선박 운항 관리 측면에서 매우 유용하게 사용할 수 있을 것으로 판단된다. 또한 2023년부터 시행되는 탄소집약도지수(Carbon Intensity Indicator; CII) 규제를 만족시키기 위해 선박의 속도를 산출하는 분야에도 충분히 활용될 수 있을 것으로 사료된다.

본 연구에서는 일반항해구간만의 데이터를 활용하여 연료소모량 예측 모델을 수립하였다는 한계가 있다. 따라서 향후 연구에서는 실선 데이터를 입출항구간, 일반항해구간, 황천항해 구간으로 구분하여 연료소모량 예측 모델의 범용성을 확보하고 정확도를 더욱 향상 시킬 예정이다. 또한 유사한 선박 크기를 가진 LNG, 메탄올 등의 친환경 선박 실선 데이터를 수집하여 본 연구와 연료소모량 비교를 통해 친환경 선박과 HFO(Heavy Fuel Oil) 선박의 에너지 절감을 비교하는 연구를 추가로 진행하고자 한다.

후 기

이 논문은 2023년도 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구이다. (20220469, 해양수산 내수시장활성화사업)

References

- [1] Chen, T. and Guestrin, C.(2016), "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.
- [2] Greenwell, B. M.(2017). "pdp:An R package for constructing partial dependence plots." The R Journal, Vol. 9, No. 1, pp. 421-436.
- [3] Hannes, J. and Linda, S.(2015), "Increased energy efficiency in short sea shipping through decreased time in port", Transportation Research Part A: Policy and Practice, Vol. 71, pp. 167-178.
- [4] IMO(2023) "Amendments to MARPOL Annex VI and the NOx Technical Code 2008 with regard to the requirement for in-use NOx abatement technology", MEPC.377(80).
- [5] Kim, Y. R., Kim, G. J. and Park, J. B.(2019), "A study on the Prediction of Fuel Consumption of a Ship Using the Principal Component Analysis", Journal of Korean Navigation and Port Research, Vol. 43, No. 6, pp. 335-343.
- [6] Lloyd, S.(1953), "A Value for n-person Games" , Contributions to the Theory of Games II, pp. 307-317
- [7] Lipton, Z. C.(2018), "The Mythos of Model Interpretability", Communications of the ACM, Vol. 61, No. 10, pp. 36-43.
- [8] Lundberg, S. M. and Lee, S. I.(2017), "A Unified Approach to Interpreting Model Predictions" Advances in Neural Information Processing Systems, Vol. 30, pp. 765-774.
- [9] Moore, D. S. and Kirkland, S.(2007), "The Basic Practice of Statistics", Vol. 2, WH Freeman Press.
- [10] Mengnan, D., Ninghao, L. and Xia, Hu.(2019), "Techniques for Interpretable Machine Learning" Communications of the ACM, Vol. 63, No. 1, pp. 68-77.
- [11] Opitz, D. and Maclin, R.(1999), "Popular Ensemble Methods: An Empirical Study." Journal of Artificial Intelligence Research, Vol. 11, pp. 169-198.
- [12] Shin, D. W., Yang, C. S. and Jeon, H. G.(2021), "Real Ship Experimental Analysis on the Influence of Fuel Oil Consumption Considering Ship Speed and Ocean Environment." Journal of Coastal Disaster Prevention, Vol. 8, No. 2, pp. 99-110.
- [13] Stopford, M.(2009), "Maritime Economics. 3rd ed.", Oxford:Routledge.
- [14] T, Uyanık, Ç, Karatuğ. and Y, Arslanoğlu.(2020), "Machine Learning Approach to Ship Fuel Consumption: A Case of Container Vessel." Transportation Research Part D: Transport and Environment, Vol. 84, 102389.
- [15] Tjoa, E. and Guan, C.(2020). "A survey on explainable artificial intelligence (xai): Toward medical xai." IEEE transactions on neural networks and learning systems, Vol. 2, No. 11, pp. 4793-4813.
- [16] Yoon, Y. J.(2005), "Marine Navigation." Haean Publishing Company, pp. 1-20.

Received 17 July 2023

Revised 21 July 2023

Accepted 04 August 2023