



공 학 석 사 학 위 논 문

# 설명가능한 인공지능(XAI)을 활용한 선박 연료 소모량 예측



산 업 및 데 이 터 공 학 과

김 현 주

# 공 학 석 사 학 위 논 문

# 설명가능한 인공지능(XAI)을 활용한 선박 연료 소모량 예측

지도교수 이 지 환

이 논문을 공학석사 학위논문으로 제출함.

2023년 8월

CH

부경대학교대학원

산 업 및 데 이 터 공 학 과

김 현 주

# 김현주의 공학석사 학위논문을 인준함.

2023년 8월 18월



표 목차
그림 목차
Abstract
I. 서론 ···································
1. 연구 배경
2. 연구 목표 및 내용
S m
II. 관련 규제 및 선행 연구 5
1. 국제해사기구(IMO)의 환경 규제 동향
1.1 운항선 효율지수(EEXI)
1.2 탄소집약도지수(CII) ······ 9
2. 선행 연구 고찰
2.1 머신러닝 기반의 선박 연료소비량 예측 관련 연구
2.2 XAI 기반의 선박 연료소비량 분석 관련 연구
Ⅲ. 선박 연료소모량 예측
1. 연구 방법 소개
2. 데이터 설명
3. 데이터 정제 및 전처리
3.1 데이터 정제
3.2 데이터 전처리



# 표 목차

Table 1.1 해운분야의 CO2 배출량 ···································
Table 1.2 2019년 기준값 대비 CII 감축계수
Table 3.1 Vessel Specification    21
Table 3.2 Original Feature    22
Table 3.3 conversion of angles    27
Table 3.4 Transformed Feature range   27
Table 3.5 final selected variables
Table 3.6 Feature correlation result   32
Table 3.7 The performance evaluation of the model using the training dataset. 37
Table 3.8 The performance evaluation of the model using the test datase 37
Table 3.9 Summarizing the relationship between the prediction
model and features through the SHAP value plot
Table 3.10 Comparison of shap value    48

# 그림 목차

Figure 1.1 국제해사기구(IMO) 선박 온실가스배출 감축전략 [4] 7
Figure 2.1 CII 등급 및 범위 산정 방법
Figure 2.2 Directional Feature Transformation
Figure 2.3 SHAP Summary Plot. 17
Figure 2.4 SHAP Feature Importance Ranking Heatmap
Figure 3.1 Flow chart of the study 20
Figure 3.2 Vessel Specification 21
Figure 3.3 Course of a vessel 22
Figure 3.4 LEEWAY and TIDEWAY 25
Figure 3.5 SWELL_WAVE_ANGLE
Figure 3.6 Vessel Draft 27
Figure 3.7 Feature distribution 30
Figure 3.8 Feature correlation matrix 31
Figure 3.9 The performance metrics of the model evaluated on the
training dataset
Figure 3.10 The performance metrics of the model evaluated on
the test dataset 38
Figure 3.11 mean absolute SHAP value 39
Figure 3.12 SHAP values(impact on model output)
Figure 3.13 Dependence plot of SPEED_VG and REL_WIND_SPEED $42$
Figure 3.14 Dependence plot of SPEED_VG and SWELL_WAVE_PERIOD 42
Figure 3.15 Dependence plot of SPEED_VG and SWELL_WAVE_HEIGHT 42

Figure 3.16 Dependence plot of REL_WIND_SPEED and AvgDraft ····· 42
Figure 3.17 Dependence plot of REL_WIND_SPEED and AvgDraft ····· 42
Figure 3.18 Dependence plot of AvgDraft and Trim 42
Figure 3.19 Dependence plot of AvgDraft and WIND_WAVE_PERIOD 43
Figure 3.20 Dependence plot of AvgDraft and SPEED_VG 43
Figure 3.21 Dependence plot of SPEED_VG 43
Figure 3.22 PDP plot of SPEED_VG
Figure 3.23 Dependence plot of REL_WIND_SPEED 44
Figure 3.24 PDP plot of REL_WIND_SPEED 44
Figure 3.25 Dependence plot of AvgDraft
Figure 3.26 PDP plot of AvgDraft 44
Figure 3.27 ME1_FOC Distribution 45
Figure 3.28 ME1_FOC Plot 46
Figure 3.29 Map showing observations with high foc (red dots) 46
Figure 3.30 Force Plot with ME1_FOC greater than 2750 47
Figure 3.31 Mean of SHAP Value comparison between All and South africa. 49
Figure 3.32 SHAP Force Plot 49
Figure 3.33 SHAP Waterfall Plot 50
Figure 3.34 Plots each feature in South Africa section (1)
Figure 3.35 Plots each feature in South Africa section (2)

#### Prediction of vessel fuel consumption using explainable artificial intelligence (XAI)

#### Hyun Ju Kim

Department of Industrial and Data Engineering, The Graduate School, Pukyong National University

## Abstract

This study proposes a prediction model for fuel consumption in ships using XGBoost and SHapley Additive exPlanation (SHAP) to explain the predicted values. Previous studies have relied solely on operational data from ships to develop prediction models, neglecting the incorporation of external weather data. However, recently, a method has been applied to increase accuracy by utilizing both operational and external weather data. Nonetheless, the reliability of the prediction results and the variables used in the prediction model implementation remained unexplained. To address these issues, XGBoost and SHAP were used in this study to develop the prediction model.

This study provides an introduction to the research background, scope, relevant regulations, and previous studies, as well as the research methodology. It also explains the data cleaning method for ships and verifies the prediction model's results. Additionally, it covers XAI-related theories and the prediction model for fuel consumption in ships using XGBoost, as well as the SHAP-based method for explaining variable influence. Finally, it discusses the final results of this research and proposes future research directions.

# I. 서 론

#### 1.1 연구 배경

2020년 세계 선박연료유 소모량은 2억 6천만톤이며, 벙커유는 전체 선박연 료의 95% 이상 차지하고 있다 [1]. 벙커유를 사용하는 선박은 주기관을 작동 시켜 추진력을 얻고 이를 통해 발전기를 돌려 선내의 각종 동력기계와 조명 등에 전기를 공급한다. 이러한 과정에서 선박에서는 많은 연료를 소비하며 많은 *CO*,를 배출하고 있다.

최근 전 세계적인 탄소중립 정책에 따라 국제해사기구(International Maritime Organization; IMO)는 2050년까지 온실가스 감축 비중을 2008년 대 비 50% 감축하기 위해 다양한 규제를 시행하고 있다.

국제해사기구(International Maritime Organization; IMO)에서는 해운 분야 탄소중립을 실현하기 위하여 에너지효율설계지수(Energy Efficiency Design Index; EEDI), 에너지효율운항지수(Energy Efficiency Operational Index;EEOI), 탄소집약도지수(Carbon Intensity Indicator; CII)와 같은 여러 환경 규제를 도 입하여 선박의 배기가스 배출량을 줄이기 위한 노력을 하고 있다. 특히, 탄소 집약도지수(Carbon Intensity Indicator; CII)는 선박의 운송 작업량 당 *CO*<sub>2</sub> 배 출량 평균을 나타는 지표로 2023년부터 기존의 국제 운항선을 대상으로 연 간 탄소배출집약도 보고를 의무화하였으며 규제를 미충족 선박에는 운항을 중단하도록 강력히 규제하고 있다.

탄소집약도지수(Carbon Intensity Indicator; CII)는 선박에서 수집되는 데이터 를 기반으로 산정되며 선박의 총톤수(Gross Tonnage; GT),연료유 종류, 연간 연료 사용량 데이터에 따라 산정된다. 탄소집약도지수(Carbon Intensity

- 1 -

Indicator; CII)에 사용되는 데이터 중에서 총톤수(Gross Tonnage; GT)와 연료 유 종류는 선박 건조와 동시에 정해지는 변동이 없는 값이다. 즉, 탄소집약도 지수(Carbon Intensity Indicator; CII)를 충족하기 위해서는 연간 연료유 소모량 을 줄이는 것이 가장 중요하다. 따라서, 정확한 연료 소비량 예측을 통한 효 율적인 선박의 관리는 환경 규제 충족을 위한 핵심 요소라고 할 수 있다.

국제적인 환경 규제뿐만 아니라 선박의 연료 비용은 전체 운항 경비의 절 반 이상을 차지하기에 해운사와 관련 업계에서도 프로펠러 효율에 따른 연 료 소비량 저감, 실선 운항 데이터를 활용한 연료 소비량 예측, 선박 도료 기술을 활용한 연료 소비량 저감 등 다양한 방식으로 선박의 연료 소모량을 예측하고 저감시키기 위한 연구가 다수 수행되어왔다. 최근에는 선박과 육상 간의 통신환경이 좋아져 실시간 선박 운항 데이터 수집이 가능해짐에 따라 빅데이터를 통한 선박 연료 소비량 예측 연구가 활발히 진행되고 있다.

## 1.2 연구 목표 및 내용

선박의 연료 소비량 예측을 위하여 수집된 선박의 빅데이터를 활용한 다 양한 연구들이 수행되었다. 기존의 많은 연구에서는 예측 불가한 선박의 외부 기상 환경 데이터를 제외한 후 선박의 운항 데이터만 활용한 예측 모 델을 구현하였다. 이러한 경우 모델의 과적합 문제가 많이 발생하였기 때 문에 최근에는 선박의 운항 데이터와 외부 기상 환경 데이터를 동시에 활 용하거나 외부의 기상 환경 데이터만을 활용하여 연료소모량 예측 모델을 구현하고 있다. 하지만 다양한 변수와 다양한 머신러닝, 딥러닝 기법을 활 용하여 예측 모델을 구현하여 정확도를 향상 시켰으나 모델 내부에서 일어 나는 의사 결정과 예측 모델 구현에 활용된 변수와 예측 모델의 결과에 대 한 신뢰성을 설명할 수 없었다. 즉, 정확한 연로소모량 예측 모델을 만들었 지만 변수로 사용된 개별 요소별로 예측 결과에 어떠한 영향을 미치는지는 확인할 수 없었다.

따라서 본 논문에서는 XGBoost을 사용하여 선박의 메인엔진 연료소모량 예 측 모델을 수립하고 모델에서 어떻게 예측값이 도출되었는지를 SHapley Additive exPlanation(SHAP)을 통하여 설명하는 예측 모델을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 1장은 연구 배경과 연구 내용 및 범위에 대하여 설명하였다, 2장에서는 정확한 선박 연료소모량 예측과 설명이 필 요한 관련 규제 및 선행 연구에 대하여 설명하였다. 3장에서는 선박 연료 소모량 예측하기 위한 본 논문의 연구방법에 대하여 소개하였다. 연구 대 상 선박에 대한 설명과 연구에 활용되는 데이터를 소개하고, 선박에서는 통신, 내부 센서 결함, 외부 환경적 문제 등로 데이터의 결측값, 이상값이 다수 발생하기에 진행한 데이터 정제 방법에 대하여 설명하였으며 예측 모 델에 대한 설명력을 높이기 위하여 변수들을 조합한 새로운 변수를 생성한 전처리 방법에 대하여 설명하였다. 다음으로 본 논문에서 사용하는 설명가 능한 인공지능(Explanable Artificial Intelligence; XAI) 관련 이론을 고찰하 였으며, 데이터 분석 실험을 통해 XGboost를 적용한 선박의 연료소모량 예 측 모델에 대하여 설명하였으며 성능 지표를 활용한 예측 모델 검증 결과 를 설명하였다. 그리고 해당 모델을 SHAP을 통하여 예측 모델에 각 변수 가 어떤 영향을 미쳤는지를 SHapley Additive exPlanation(SHAP)통하여 설 명하였다. 마지막으로 4장에서는 본 연구의 최종 결과와 향후 연구에 대하 여 설명하였다.



# Ⅱ. 관련 규제 및 선행 연구

## 2.1 국제해사기구(IMO)의 환경 규제 동향

국제해사기구(International Maritime Organization; IMO)는 해양오염방지협약 (International Conventionfor the Prevention of Marine Pollution from ships; MARPOL)에 의해 국제 해운분야의 온실가스 배출에 대한 규제가 이루어지고 있다. 해양환경보호위원회(Maritime Environment Protection Committee; MEPC)에서 국제 해운선박의 *CO*<sub>2</sub> 배출량에 대한 연구인 1차 선박온실가스 배 출량 분석연구(Greenhouse Gas Study; GHG Study)를 수행하였고, 현재 4차까 지 보고서가 발간되었다. 이에 따르면 해운분야에서 배출되는 *CO*<sub>2</sub> 배출량이 2012년 9억 6,100만 톤에서 2018년 10억 5,600만 톤으로 증가하였음을 알 수 있다 [2].

Table 1.1 해운분야의 *CO*<sub>2</sub> 배출량

단위 : 백만톤

연도	전체	해운	비율(%)						
2012	34,793	961	2.76						
2013	34,959	956	2.74						
2014	35,224	963	2.74						
2015	35,238	990	2.81						
2016	35,379	1,026	2.90						
2017	35,810	1,063	2.97						
2018	36,572	1,056	2.89						

국제해사기구(International Maritime Organization; IMO)는 GHG(Greenhouse Gas Study; GHG Study)에서 신조 선박 EEDI의 추가 단계 시행을 통한 탄소집 약도 감소, 국제 해운 탄소 집약도 감소, 국제 해운 온실가스 배출량 정점 달 성 및 감소와 같은 3가지의 감축 목표를 제시하고 있다 [3]. 특히 국제 해운 탄소 집약도 감소를 위하여 단기, 중기, 장기 목표를 세우고 온실가스 감축을 위한 방법을 제시하였다. 단기적으로 2023년도는 2008년도 대비 30%, 중기적으로 2030년도는 40%, 장기적으로 2050년도는 50% 감축을 목표로 하고 있다. 단기적으로 IMO MEPC는 온실가스 감축을 위해 크게 기술 적, 운항적, 시장기반적 솔루션 개발을 위한 방안을 마련하고 있다. 기술적 조 치로는 선박에너지효율설계지수(Energy Efficiency Design Index; EEDI)가 있 고, 운항적 조치로는 선박에너지효율관리계획(Ship Energy Management Plan; SEEMP)과 선박운항에너지효율지수(Energy Efficiency Operationma Indicator; EEOI)가 있다.

2021년 6월 IMO MEPC에서 MARPOL이 개정됨에 따라 현존선에 대한 온실 가스 감축을 위한 규제가 적용된다. 기술적 조치로는 선박에너지효율지수 (Energy Efficiency Existing Index; EEXI) 제도, 운항적 조치로는 선박탄소집약 도(Carbon Intensity Indicator; CII) 제도가 추가되었다.

01 II

1



## 2.1.1 운항선효율지수(Energy Efficiency Existing Ship Index; EEXI)

에너지효율설계지수(EEDI, Energy Efficiency Design Index)는 신조선에 적 용되는 대표적인 기술적 조치 중 하나이며 2013년 이후 건조되는 선박에 대해 설계단계에서부터 적용되는 규제이다. 선박의 설계과정에서 1톤의 화 물을 1해리 운송하는데 배출되는 *CO*<sub>2</sub> 양을 기관출력, 재화중량톤수 등 선 박의 제원을 활용해 사전적으로 계산 및 지수화한 값이며. 2013년 1월 1일 이후 건조된 선박에 적용된다. 선종 및 화물량별로 에너지 효율지수의 기 준을 각각 적용하고 있다.

에너지효율설계지수(EEDI, Energy Efficiency Design Index)값은 식 2.1과

같이 나타낼 수 있다 [5].

$$EEDI = \frac{(\prod_{j=1}^{M} f_j)(\sum_{\substack{i=1\\nPTI}}^{nME} P_{ME(i)} \bullet C_{FME(i)} \bullet SFC_{ME(i)}) + (P_{AE}C_{FAE}SFC_{AE})}{(\prod_{j=1}^{M} f_j \sum_{i=1}^{nPTI} P_{PTI(i)} - \sum_{i=1}^{\neq ff} f_{eff(i)} P_{AEeff(i)})C_{FAE}SFC_{AE}} - (\sum_{i=1}^{\neq ff} f_{eff(i)} P_{eff(i)} \bullet C_{FME} \bullet SFC_{ME})}{f_j Capacity V_{ref} f_w}$$

$$(2.1)$$

EEDI는 설계속력을 반영하여 CO2를 계산하며 EEXI는 운항속력을 사용하 여 CO2를 계산한다. 설계속력이란 일반적으로 엔진 MCR(Maximum Continuous Rating)의 75~83%정도로 계산된다 [6], [7]. 이는 선사의 지침, 여러 환경적 조건에 따라 선박별 속도가 상이하게 운항되는 운항속력과는 차이 가 있다. 운항선효율지수(Energy Efficiency Existing Ship Index; EEXI)값은 식 2.2과 같이 나타낼 수 있다 [5].

$$EEXI = \frac{(\prod_{j=1}^{M} f_j)(\sum_{i=1}^{nME} P_{ME(i)} \bullet C_{FME(i)} \bullet SFC_{ME(i)}) + (P_{AE}C_{FAE}SFC_{AE})}{Capacity V_{ref}f_i f_c f_t f_w f_m}$$
(2.2)

fi Capacity .Vref fw: Efficiency Technologies fi Capacity .Vref fw: Transport work PME: Main Engine Power (kW) PAE: Auxiliary Engine Power (kW) SFC: Specific fuel consumption (g/kW)

C: Fuel to CO2 factor (g CO2/ g Fuel) (nearly 3)

Capacity: for Cargo ships DWT, for Passenger ships GT

Vref: Reference speed (nm/h)

fi: Correction factor for capacity

fw: Correction factor for performance in real weather

fi: Correction factor for efficiency

설계속력과 운항속력이라는 차이점 이외에 EEXI 규제는 EEDI 규제와 동 일하게 1톤의 화물을 1해리 운송하는데 배출되는 *CO*<sub>2</sub> 양을 측정한다. 400톤 이상의 선박은 운항선효율지수(Energy Efficiency Existing Ship Index; EEXI)라는 규제를 2023년부터 만족해야한다. 따라서, 2023년부터는 규정에 따라 현존 운항 선박에 대해서 EEXI허용값을 만족시켜야하며 허용값 미충족 선박의 경우에는 SHaPoli(Shaft Power Limitation) 또는 EPL(Engine Power Limitation) 장비를 의무 설치해야한다.

# 2.1.2 탄소집약도지수(Carbon Intensity Indicator; CII)

총톤수 5,000톤 이상 국제선에 적용하는 규제로 선박의 운항과정에서 1 톤의 화물을 1 해리 운송하는데 배출되는 CO2 양을 연료사용량, 운항거리 등 선박의 운항정보를 활용하여 사후적으로 계산 및 지수화한 값이다. 2023년부터 운행중인 선박은 매년 일정 수준의 온실가스 배출량을 감축시 켜야 한다. 온실가스 배출량 감축에 대하여 검사 결과에 목표 달성도에 따 라 등급(A~E)을 부여하고, 해당 기간의 선박에 요구되는 CII허용값 (Required) 대비 CII달성값(Attained) 정도에 따라서 A~E등급이 부여된다.

. 단일 연도 E등급 또는 3년 연속 D등급을 받은 선박은 SEEMP의 시정조치 계획을 승인받은 후 운항이 허용된다.

CII 등급을 도출하기 위한 CII 달성값은 식 2.1과 같이 나타낼 수 있다 [3].

$$CII 달성값(Attained CII) = \frac{\sum (FC_j \times C_{f_j})}{Capacity \times Total Distance}$$
(2.1)

이때 FC<sub>j</sub>는 유종별 연간 연료 사용량, C<sub>fj</sub>는 연료사용량을 CO<sub>2</sub>배출량으로 변환하는 유종별 변환계수로, EEDI산정 가이드라인에 명시된 값이다. Capacity는 선박 용량이(총 톤수 또는 재화중량톤수)며 Total Distance는 1 년간 해당 선박이 운항한 총 거리를 의미한다 [3].

CII 허용값은 식 2.3과 같이 나타낼 수 있다. CII 허용값을 산정하기 위해 서는 식 2.2와 같은 기준값이 필요하다 [3].

CII기준값(reference CII) = a Capacity<sup>-</sup>

(2.2)

이때 a와 c는 2019년 IMO DCS데이터를 활용하여 산출한 개별 선박의 용 량과 CII 달성값을 기반으로 도출한 매개변수로 선종 및 크기별로 정해진 값으로 EEDI산정 가이드라인에서 제공하고 있다 [3].

Н

$$CII \stackrel{\text{de}}{\to} \stackrel{\text{de}}{\to} (Required CII) = (1 - \frac{Z}{100}) \times CII_{ref}$$
 (2.3)

이때 Z는 2019년 기준값 대비 연도별 CII 감축계수로 모든 선박에 대해

동일한 감축계수를 적용하며 마찬가지로 EEDI산정 가이드라인에서 제공하 며, 2023년부터 적용되며 매년 2%씩 추가 감축을 해야한다.

Table 1.2 2019년 기준값 대비 CII 감축계수

연도	2023	2024	2025	2026
감축계수	5%	7%	9%	11%

마지막으로 CII등급은 A~E까지 있으며, 각 등급별 범위는 CII허용값을 기 반으로 식 2.4-2.7과 같이 각 등급별로 산정한다 [3].



Figure 3.1 CII 등급 및 범위 산정 방법

Superior boundary = $\exp(d_1) \times required CII$	(2.3)
Lower boundary = $\exp(d_2) \times required CII$	(2.4)
Upper boundary = $\exp(d_3) \times required CII$	(2.5)

 $Superior boundary = \exp(d_4) \times required CII$ (2.7)

exp(d)는 CII허용값으로부터 벗어나는 방향과 거리를 나타내는 벡터의 지 수로서 선종 및 크기별 값은 EEDI 가이드라인에서 제시한다.

## 2.2 선행 연구 고찰

선박의 통신환경이 좋아짐에 따라 실시간으로 선박 데이터 수집이 가능해 졌다. 따라서, 최근에는 연료소모량을 예측하는 연구는 1)통계적 분석, 2)기 계학습 기법을 많이 사용한다. 연료소모량 예측 모델 중심으로 한 선행 연구 의 방법, 평가지표에에 대하여 조사하였고, 선행 연구의 한계점과 본 논문에 서 이를 보완한 방법에 대하여 작성하였다.

## 2.2.1 머신러닝 기반의 선박 연료소비량 예측 관련 연구

해당 논문은 1척의 컨테이너 선박 데이터(328m×46 m×9.7m container ship)를 기반으로 다양한 통계적 기계학습 기법을 사용해 예측 모델을 생 성하였다. 예측 모델에 사용된 데이터와 수집방식, 전처리 방식, 학습 방 법, 모델 평가 방법에 대하여 아래와 같이 정리하였다.[8]

 데이터 수집 및 전처리 : 총 75개의 매개변수, 724개 데이터를 사용하였으며, 센서 오작동으로 인한 에러, 휴먼에러와 같은 비정상적인 데이터는 2 엔진 제조사 operating profile 기준으로 처리하였다. 각각의 데이터 수집방 식은 아래와 같다.[8]

- Noon Report(정오 보고서) : Shaft power, rpm, and torque, M/E rpm, and fuel oil consumption

- Engine Log Book(엔진 로그북) : M/E power, scavenging air receiver temperature, scavenging air pressure, M/E cooling water temperature and pressure, M/E thrust power
- 이 외 센서류 : M/E fuel flow counter, exhaust outlet temperature of cylinders

2) 상관관계 분석 : Pearson 방법을 통해 변수간의 관계 분석을 하였으며 아래의 변수들과 연료소모량간의 관계를 분석하여, main engine cylinder values, scavenge air parameters, cooling water values, shaft parameter indicators 가 가장 높은 상관관계를 가짐을 확인하였다.[8]

3) 학습 방법 : 전체 데이터 724개 중 학습용 데이터는 66%(485개), 테스트 데이터 33%(239개)로 나누어 회귀 방법(Multiple linear regression, Ridge & LASSO regression, Kernel Ridge regression, Bayesian Ridge regression, Support Vector regression, Multi-Layer Perceptron regression, Decision tree regression, Random Forest regression, Ada Boost regression, Gradient Boosting regression, Hist Gradient Boost regression, K-Nearest Neighbors) 총 12개의 모델을 사용하여 학습하였다.[8]

4) 평가 방법 : 테스트 데이터 33%(239개)을 K-fold cross validation 기법을 사용하여 모델을 평가하였으며, 예측의 정확도에 대한 성능지표로는 RSME(Root mean square error), MAE(Mean absolute error), R<sup>2</sup>(Coefficient of determination)를 사용하여 모델 별 예측 정확도를 평가 및 비교하였다.[8]

5) 평가 결과 : R<sup>2</sup>(Coefficient of determination) 지표로 모델 예측성을 평

가한 결과 Bayesian Ridge, Kernel Ridge, Multiple Linear and Ridge Regressions가 가장 높은 예측정확도를 보이는 것으로 평가되었다. [8]

- 6) 연구의 차별성 확보 방안 : 해당 참고 논문에서는 엔진의 주기관에서 채취한 샘플을 바탕으로 연료유 소모량 추정 연구를 수행하였으
   나 아래와 같은 한계점을 분석하였다.
- 연료소모량을 예측하는데 외부의 해상 환경 데이터(선박 속도, 풍속, 흘수, 트림)를 고려하지 않았다.
- 1척의 컨테이너 선박을 대상으로 총 724개의 데이터만을 사용하였다.
- 변수 간 상관분석이 이루어져 각 변수가 서로에 미치는 영향값을
   결정하여 연료 소비량 예측 모델에 사용되었으나, 상관관계가 높은 변수
   끼리의 독립성을 보장하지 못했다.

따라서 본 연구에서는 엔진의 주기관 데이터는 예측의 변수로 활용하지않 으며 외부의 환경 데이터만을 이용하였다. 선박의 메인엔진의 파워, rpm 등은 선박의 연료소모와 상관관계가 아주 높은 변수들이다. 엔진 데이터를 활용하지 않음으로서 변수간의 상관관계가 높은 변수를 제거할 수 있었으 며 선박의 연료소모량을 예측하는 연구의 차별성을 확보하였다.

## 2.2.2 XAI기반의 선박 연료소비량 분석 관련 연구

해당 논문은 18만톤급의 벌크선 데이터를 기반으로 Catboost 회귀 모델을 사용하여 선박의 축마력해 예측 모델을 생성하며 생성된 모델에 대해 SHapley Additive exPlanations (SHAP) 방법을 적용하여 각 예측 변수가 예 측 결과에 미치는 기여정도를 설명하였다.[9]  데이터 수집 : 기존의 데이터셋 151개 feature 중에서 축마력 예측을 위 한 17개의 feature만을 사용하였으며 16개의 다른 항로를 가진 벌크선의 16개월간의 선박 엔진 센서 데이터와 Noon Report를 활용한 데이터셋을 활용하였다.[9]

2) 데이터 전처리 : 각 feature에서 -9999로 나타나는 에러 값을 포함하는 데이터 instance는 모두 제거되었다. 그 후, 일부 feature 값은 합리적인 기 준으로 필터링되었다. 예를 들어, 전체 파고 높이는 6m를 초과하지 않도록 설정되었고, 지상 속도는 10노트 이상으로 설정되었다.[9]

3) Feature transformation : SHIP HEADING, RUDDER ANGLE,
 REL\_WIND\_ANGLE 등을 0°에서 360°사이의 벡터값을 가진 degree 변수
 들을 SHIP HEADING과의 차로 계산하여 새로운 변수를 생성하였다. 또한,
 선박의 선수, 선미의 Draft 값의 산술평균으로 새로운 변수를 생성하였다.[9]



Figure 2.2 Directional Feature Transformation [9]

4) 학습 및 평가 방법 : . Catboost 회귀 모델을 사용하여 선박 메인엔진의 축마력을 예측하였다. 분석은 1차적으로 16개의 항차에 관계없이 모든 데 이터를 사용한 50,798개의 인스턴스가 모델에 활용되었고, R-squared, RMSE, MAE 및 MAPE와 같은 4가지 지표를 계산하여 예측 모델을 평가하였다. 또 한 모델을 최적화하기 위해 다른 항해별로 교차 검증을 진행하였다. [9]

5) 평가 결과 : R-squared 평가결과 16개의 항차 평균 0.93의 정확도를 가짐 을 확인할 수 있었으며 일부 항차가 R-squared 값이 평균에 비해 0.8정도로 낮았지만 RMSE, MAE 및 MAPE에 의해 나타나는 오차율은 평균 항해 오차율 보다 크지 않았으며 이는 전체 항차의 평균 오차율도 초과하지 않았다. [9]

6) 모델 해석 : Figure 2.3과 같이 SPEED\_VG가 축마력 예측에 가장 영향을 미치는 가장 중요한 특성으로 나타났으며, 그 다음으로 REL\_WIND\_SPEED 와 AvgDraft가 축마력 예측에 영향을 주는 순위로 나타났다. 또한 세부적 으로 SPEED\_VG의 SHAP Dependence Plot에서 SPEED\_VG값이 높을수록 더욱 높은 Shaply Value값을 가지는 것을 확인하였으며, 값이 16 knots 이 하로 떨어지면, SPEED\_VG는 예측 결과에 부정적인 영향을 미치거나 축력 계산에 부정적인 영향을 준다는 것을 확인하였다. REL\_WIND\_SPEED와 AvgDraft에서도 동일한 방식으로 모델을 해석하였으며, 상대풍속 (REL\_WIND\_SPEED)의 경우, 10.5 m/s 이상일 때 예측 결과에 긍정적인 영 향을 미치며 AvgDraft가 8 이상일 때 예측 결과를 증가시키는 경향을 확인 하였다. [9]



마지막으로, SHAP에 의해 계산된 각 feature의 Shapley 값의 평균에 기반 하여 모든 데이터 및 각 항해 번호별 분석의 feature 중요도 순위를 확인 하였다. 평균 Shapley 값에 따라 가장 영향력이 큰 상위 세 가지 feature는 모델이 다른 항해 번호를 분석할 때 feature 중요도 순위가 변동되는 것을 확인하였으나 대부분 항차별로 SPEED\_VG와 REL\_WIND\_SPEED는 상위권 이며 AvgDraft는 항차별로 중요도가 변동됨을 Figure 2.4와 같이 확인할 수 있었다. [9]

Feature Importance	Overall								SHA	P Fea	ture l	mpor	tance	Rank	per V	/oyage	e Nur	nber							
Rank Da	Data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
SPEED_VG	#01	#01	#01	#03	#01	#02	#01	¢01	#01	#01	#01	#01	#01	#01	#01	#01	#01	#01	#01	#02	#01	#01	#01	#01	#01
REL_WIND_SPEED	#02	#03	#03	#02	#03	#01	#05	#02	#02	#03	#04	#02	#02	#02	#06	#02	#02	#02	#02	#06	#02	#02	#04	#02	#04
AvgDraft	#03	#09	#07	#01	<i>‡</i> 02	#10	#10	#05	#07	#04	#02	#04	#03	#09	#04	#03	#04	#05	#07	#13	#03	#03	#02	#03	#03
SEA_SURFACE_TEMP	#04	#12	#14	#13	#11	#07	#08	#04	#05	#13	#13	#10	#09	#03	#05	#04	#05	#08	#10	#05	#07	#08	#05	#09	#07
WIND_SPEED	#05	#07	#02	#07	#10	#08	#07	#03	#08	#08	#03	#09	#06	#07	#08	#07	#06	#04	#12	#08	#10	#06	#09	#12	#11
TOTAL_WAVE_HEIGHT	#06	#02	#06	#09	#09	#09	#09	#06	#10	#02	#06	#08	#04	#06	#03	#06	#11	#03	#04	#04	#06	#11	#03	#08	#02
WaveD	#07	#08	#04	#06	#04	#03	#02	#10	#06	#10	#07	#07	#13	#10	#07	#14	#12	#10	#06	#01	#11	#07	#13	#05	#09
SEA_SURFACE_SALINITY	#08	#05	#08	#12	#07	#12	#03	#07	#11	#05	#05	#12	#05	#05	#02	#12	#14	#09	#03	#03	#12	#05	#07	#11	#05
RudD	#09	#06	#09	#04	#05	#04	#06	#08	#03	#06	#14	#03	#11	#04	#09	#05	#03	#06	#08	#10	#05	#04	#06	#07	#06
WATER_DEPTH	#10	#04	#11	#10	#06	#11	#12	#13	#12	#12	#11	#11	#10	#11	#11	#11	#07	#13	#05	#12	#04	#10	#11	#04	#13
CurrD	#11	#13	#10	#08	#12	#06	#11	#11	#14	#11	#08	#05	#12	#12	#13	#13	#10	#07	#09	#14	#08	#14	#10	#10	#08
CURRENT_SPEED	#12	#14	#05	#14	#14	#13	#14	#14	#13	#09	#12	#13	#14	#14	#14	#10	#09	#12	#13	#07	#14	#13	#14	#14	#14
WindD	#13	#11	#12	#05	#08	#05	#04	#09	#04	#07	#09	#06	#08	#08	#12	#08	#08	#11	#11	#09	#09	#09	#08	#06	#10
RelWD	#14	#10	#13	#11	#13	#14	#13	#12	#09	#14	#10	#14	#07	#13	#10	#09	#13	#14	#14	#11	#13	#12	#12	#13	#12

Figure 2.4 SHAP Feature Importance Ranking Heatmap [9]

7) 연구의 차별성 확보 방안 : 해당 참고 논문에서는 축 마력이 크게 나타 난 구간에 대해서는 별도의 해석을 하지 않았으며 SHIP HEADING과의 WIND, SWELL, CURRENT DIRECTION 산술차이를 통해 구하였기 때문에 선박의 실제 운항 환경과는 차이가 있다.

따라서 본 연구에서는 WIND, SWELL, CURRENT DIRECTION을 모두 예각 처리하였으며 선박 연료소모량이 크게 나타나는 구간과 전체 데이터셋과의 SHAP Value을 추가로 비교분석하였다.

# Ⅲ. 선박 연료 소모량 예측

## 3.1 연구 방법 소개

본 논문에서는 Fig. 3.1 연구 흐름과 같이 연구를 진행한다. 수집된 선박 의 Raw Data 정제를 포함하여 유의미한 결과 도출을 위하여 데이터 표준 화, 변수들의 조합을 통한 새로운 변수 생성과 같은 전처리 작업을 진행하 였다. 다음으로 예측 모델 생성을 위하여 데이터를 학습과 테스트를 위하 여 8대 2의 비율로 분리하였으며, XGboost 모델을 적용하여 선박 연료소모 량 예측 모델을 개발하였다. 다음으로 예측 모델 평가를 통하여 모델의 정 확도를 확인하였다. 마지막으로 SHAP values를 이용하여 개발한 예측 모델 에 사용된 변수들이 모델의 예측에 어떤 영향을 미치는지를 설명하였다.





Figure 3.1 Flow chart of the study

# 3.2 데이터 설명

본 연구에서는 사용된 대상 선박의 정보는 Table 3.1과 같다. 대상 선박 은 160,0000 톤급의 대형 화물 선박이다. 대형 화물 선박의 경우에는 선박 연료 소모량이 운영 비용에 직접적인 영향을 미치기에 선사에서 관심이 많 은 선종 중 하나이며 최근 국제적인 선박의 환경 규제에서 5,000톤급 이상 의 국제 원양 선박에 대해서는 연료소모량이 중요 지표로 작용하는 규제 등급을 산정하는 점을 고려하여 해당 선박으로 선정하였다. 또한, 화물 선 박은 주로 대서양과 태평양을 연결하는 긴 항로를 항해하는 경우가 많아 데이터 분석시에 안정적인 데이터양을 수집할 수 있다는 것을 고려하여 최 종 분석 대상 선박을 선정하였다.

Table 6.1 Vessel Specification

Specification	Measurement
Type of ship	(BULK) CARRIER
L.O.A (m)	333.07
L.B.P (m)	327
Beammoduled(m)	60
Depth moduled (m)	29.8
Gross Tonnage(ton)	160,290
Net Tonnage(ton)	56,873



Figure 3.2 Vessel Specification



Figure 3.3 Course of a vessel

대형 화물 선박에서는 위치, 속도, 엔진 상태, 기상정보, 선체 상태 데이 터 등 실시간으로 다양한 데이터가 수집된다. 본 연구에서는 이러한 선박 데이터를 10초 주기로 수집된 데이터를 10분 주기로 압축한 데이터를 활용 하여 연구를 진행하였다. 데이터는 10초단위로 수집된 10분동안의 데이터 를 평균치로 압축하여 데이터 수집 업체로부터 제공받았다.

본 연구에서는 대상 선박이 2020년 01월 01일부터 2020년 06월 30일까지 의 6개월의간의 선박 운항 데이터를 연구에 사용하였다. 변수의 총 개수는 190개이나 본 연구에서는 19개의 변수를 활용하였으며 각 Feature에 대한 설명을 Table 3.2와 같이 나타내었다,

	Feature Name	Description	Unit
1	ME1_FOC	no.1 ME Consumption	kg/h
2	SPEED_VG	Speed over ground	knots

		<b>T</b> T 1 11					
	Vertical distance between the						
3	DRAFT FORE waterline and the bottom of the hull						
		measured at the perpendicular of					
		the bow					
		Vertical distance between the					
4	DRAFT AFT	waterline and the bottom of the hull	m				
-		measured at the perpendicular of					
		the stern					
		Vertical distance between the					
5	DRAFT MID PORT	waterline and the bottom of the hull	m				
0		measured at the perpendicular of					
	ALAT	the port side (left)					
	CN	Vertical distance between the					
6	DRAFT_MID_STBD	m					
0		measured at the perpendicular of					
		thestarboard side(right)					
7	SHIP HEADING Direction of indicated by the						
		vessel's own needle					
8	COURSE OVER GROUND	Direction of which the vesselis	degree				
	101	moving					
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_DIRECTION	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	REL_WIND_SPEED	Speed of the apparent wind	m/s				
19	REL_WIND_DIR	Direction of the apparent wind	degree				

## 3.3 데이터 정제 및 전처리

#### 3.3.1 데이터 정제

선박에서는 센서 노후화, 기계적 이상, 통신 문제 등 다양한 이유로 수집 된 센서 데이터 오류가 발생한다. 본 연구에 활용된 운항 데이터는 오류나 결측치가 발생할 경우 -9999와 error와 같이 기록되며 본 연구에서는 이러 한 오류 데이터를 삭제하는 정제작업을 수행하였다.

# 3.3.2 데이터 전처리

본 연구에서는 변수들간의 조합을 통하여 새로운 변수를 생성하였다. 변 수들간의 조합을 통해 변수간의 상호작용을 파악할 수 있으며 이는 결국 예측 모델의 정확도를 높이고 예측 모델에 대한 설명성을 높이는 장점이 있다. 본 연구에서는 선박의 연료 소모량 예측에 사용되는 변수 중 어떠한 환경적인 변수가 연료 소모량에 어떤 영향을 미치는지를 설명하고자 하였다. [18]에 따르면 Leeway(바람힘에 의한 이탈운동)는 선박의 실제 이동방향과 거리감시장치 등을 이용하여 측정되는 선박의 원하는 이동방향 사이의 각 도라고 정의하며, 이를 풍압차라고 한다. 즉, Course over groud와 Ship heading의 차이를 Figure 3.4와 같이 풍압차(Leeway)라고 정의하며 새로운 변수로 추가하였다. 또한 [18]에서 Tideway(조류운동)를 선박이 흐르는 강 이나 해상에서 발생하는 조류라고하며, 이러한 조류 영향으로 선박의 이동 방향과 항로가 일치하지않는 경우를 말한다. 본 연구에서는 이러한 개념을 활용하여 유압차(Tideway)를 조류의 방향과 Ship heading의 차이로 정의하 였다.



또한, 외부의 기상조건(풍력, 해류, 조류 등)에 의한 선박 연료소모량에 미 치는 영향도를 세부적으로 설명하고자 바람, 너울, 조류 등의 방향과 SHIP HEADING과의 각 차이를 Figure 3.5와 같이 추가하였다.



본 연구에서는 선박의 연료 소모량 예측에 선박이 항해할 때 외부의 기상 조건(풍력, 해류, 조류 등)에 의한 영향도를 설명하고자 하였기에 LEEWAY, TIDEWAY, SWELL\_WAVE\_ANGLE, REL\_WIND\_ANGLE을 모두 0°~90° 사 이의 Table 3.3과 같이 예각으로 변환처리를 하였다. 이는 외부의 환경이 SHIP HEADING에 미치는 영향을 보다 더 정확하게 파악하고자 위함이다.

Table 3.3 conversion of angles

	Range	Processing	Unit
1	$90~\degree <  m angle < 180~\degree$	angle - 90°	degree
2	180 ° <angle<270 td="" °<=""><td>angle - 180 °</td><td>degree</td></angle<270>	angle - 180 °	degree
3	270 ° <angle<360 td="" °<=""><td>angle – 270 °</td><td>degree</td></angle<360>	angle – 270 °	degree

선박은 평형상태는 선박이 운항시에 복원성과 직결되는 문제이다. 선박에 서는 이러한 평형상태를 선수, 선미, 좌현, 우현으로 나누어 선체가 물에 잠긴 정도를 측정하여 유지한다. 선수홀수와 선미홀수의 차이를 트림(Trim) 이라 하며 이는 결국 선박이 앞뒤로 기울어진 정도를 의미한다. 비슷하게 좌현(PORT)과 우현(Starboard)의 차이를 힐(Heel)이라고 하며 이는 선박이 좌우로 기울어진 정도를 의미한다.

따라서 본 연구에서는 선박의 4개 측면에서의 홀수 산술평균을 새로운 변 수로 추가하였으며, 선수와 선미, 좌현과 우현의 홀수 차이를 각각 Figure 3.6과 같이 추가하였다. Table 3.4에 변수간의 조합을 통하여 새로운 변수 를 생성한 변수들의 범위를 정리하였다.



Figure 3.6 Vessel Draft

	Feature Name	count	mean	std	min	max
1	AvgDraft	16542	9.116903	5.909587	0	17.737
2	Trim	16542	9.422862	6.180069	0	32.839

3	Heel	16542	172.07788	137.85486	0	359.993
4	LEEWAY	16542	182.84588	97.866754	0.019	359.985
5	TIDEWAY	16542	170.64845	100.86274	0	359.775
6	WIND_WAVE_ANGLE	16542	14.635316	5.414836	7.61	22.715
7	SWELL_WAVE_ANGLE	16542	15.959232	5.019753	10.11	23.144
8	REL_WIND_ANGLE	16542	15.530226	5.385416	9.27	23.23

정리하면 본 연구에서는 변수들간의 조합을 통한 새로운 변수를 포함하 여 총 17개의 변수를 Table 3.5와 같이 연구에 사용하였다.

	Feature Name Description		Unit
1	ME1_FOC	no.1 ME Consumption	kg/h
2	SPEED_VG	Speed over ground	knots
3	AvgDraft	Averaged value of DRAFT_FORE DRAFT_AFT, DRAFT_MID_PORT and DRAFT_MID_STBD	m
4	Trim	Difference between DRAFT_AFT and DRAFT_FORE	m
5	Heel	Difference between DRAFT_MID_PORT and DRAFT_MID_STBD	m
6	LEEWAY	Angular difference between COURSE_OVER_GROUND and SHIP_HEADING	degree
7	TIDEWAY	Angular difference between CURRENT_DIRECTION and	degree

Table 3.5 final selected variables

		SHIP_HEADING	
		Angular difference between	
8	WIND_WAVE_ANGLE	WIND_WAVE_DIRECTION and	degree
		SHIP_HEADING	
		Angular difference between	
9	SWELL_WAVE_ANGLE	SWELL_WAVE_DIRECTION and	degree
		SHIP_HEADING	
		Angular difference between	
10	REL_WIND_ANGLE	REL_WIND_DIR and	degree
		SHIP_HEADING	
11	RUDDER_ANGLE	angle of rudder	degree
12	WIND_WAVE_HEIGHT	Significant height of wind waves	m
13	WIND_WAVE_PERIOD	Mean period of wind waves	s
14	CURRENT_SPEED	Directional movement of seawater	m/s
15	SWELL_WAVE_HEIGHT	Significant height of swell waves	m
16	SWELL_WAVE_PERIOD	Mean period of swell waves	s
17	REL_WIND_SPEED	Speed of the apparent wind	m/s

본 연구에서는 활용되는 선박의 MCR(Maximum Continuous Rating)의 75~83% 구간의 메인엔진의 RPM과 Speed를 고려하여 ME1\_RPM 50이상, SOG 10knot 이상의 선박이 안정적으로 항해하는 구간의 데이터를 활용하였 다. 또한 이러한 구간에서는 선박이 경제적이고 안정적이라는 이유로 통상적 으로 일정 방향을 유지하며 항해하는 구간이라고 볼 수 있다. 목적지에 도달 하기 위하여 방향을 바꾸는 경우를 제외한다면 SHIP HEADING 과 COURSE OF GROUND의 각의 차이가 크게 발생했다면 항해 구간에서 발생하는 외력의 영향이라고 추측해 볼 수 있기 때문이다. 변수들의 분포를 subplot을 통해 Figure 3.7과 같이 나타내었다.



Figure 3.7 Feature distribution

본 연구에서 사용한 변수들간의 상관관계 분석 결과를 Figure 3.8, Table 3.6과 같이 나타내었다. 타겟변수인 선박 메인엔진의 연료소모량(ME1\_FOC)과 SPEED\_VG(Speed Over Ground; SOG)와의 상관계수가 0.32로 ME1\_FOC에 SPEED\_VG가 가장 큰 상관성이 있는 것으로 보였다. 다음으로 선박 메인엔진

의 연료소모량(ME1\_FOC)과 REL\_WIND\_SPEED와의 상관계수가 0.23, Trim과 의 상관계수가 0.15, Heel과의 상관계수가 -0.17로 나타났다. 선박의 속도, 선 박의 상대풍속, 선박의 트림, 힐이 연료소모량에 선형관계로 상관관계를 가짐 을 확인하였다.



Figure 3.8 Feature correlation matrix

Table 3.6 Feature correlation result

	Target	SPEED_VG	REL_WIND_	Trim	II.11
	Feature	(SOG)	SPEED	1 rim	Hell
1	ME1_FOC	0.32	0.23	0.15	-0.17

데이터 전처리 단계의 마지막으로 본 연구에서 활용한 데이터의 경우에 는 선박의 여러 센서로부터 수집되어 서로 다른 단위와 범위를 가지고 있 다. 따라서 본 연구에서는 평균이 0이고 표준편차가 1인 정규분포로 변환 하는 방식인 Standard Scaler를 활용하여 데이터 표준화 작업을 수행하였 다. 이를 통해 데이터의 범위를 조정하고, 다양한 스케일의 변수를 동일한 스케일로 맞춰줌으로써 모델의 예측력을 높이고자 하였다.

Standard Scaler는 각 변수(feature)의 값들을 평균과 표준편차로 변환하는데, 이는 다음과 같은 수식으로 계산된다.

 $Z = \frac{(x-u)}{s}$ 

(3.1)

여기서 x는 원래 값, u는 변수의 평균, s는 변수의 표준편차이다.

### 3.4 XAI 관련 이론 고찰

XAI(Explainable Artificial Intelligence)는 인공지능 모델의 동작 방식을 설명 가능하도록 만드는 기술이다. XAI는 인간이 이해할 수 있는 방식으로 작동 원리와 의사 결정 과정을 설명하는 AI 시스템을 개발하는 것을 목적 으로 한다 [10]. AI는 머신러닝 모델의 예측 결과를 해석하고, 모델의 내부 동작을 분석하여 인과적인 관계나 패턴 등을 추론할 수 있다. XAI의 이론 은 크게 모델 내부 설명력과 모델 외부 설명력으로 구분할 수 있으며, 모 델 내부 설명력은 모델 자체가 설명 가능한 모델인지 여부를 평가하며, 모 델 외부 설명력은 모델을 해석하는 기술과 방법을 의미한다 [11].

모델 외부 설명력은 다양한 시각화 기술, 질문-답변 형태의 인터페이스, 자연어 처리 등의 기술로 모델을 해석하는 것을 의미하여 데이터의 특징 이나 모델의 동작 방식 등을 이해하기 쉽도록 가시화하여 제공한다. 질문-답변 형태의 인터페이스는 사용자가 모델을 직접 조작하면서 모델의 결과 를 해석할 수 있도록 하며 자연어 처리는 모델의 결과를 인간이 이해하기 쉬운 자연어 형태로 변환하여 제공한다 [12].

즉, XAI기법은 인간과 인공지능 간의 의사소통을 지원하며. 모델의 결과 가 설득력있게 전달하고, 모델의 내부 동작이 명확하게 이해시킬 수 있도 록 활용되는 방식이다 [12].

SHAP이 등장하기 이전의 Feature Attribution 방법들은 대부분 Local Interpretable Model-Agnostic Explanations (LIME)와 Integrated Gradients 이었다. LIME은 모델이 어떻게 작동하는지를 이해하기 위해 개별 데이터 포인트 주변에서 모델을 해석할 수 있는 기법이다. LIME은 모델을 해석 가 능한 모델로 근사화하여 해당 모델을 해석하며 LIME은 해당 데이터 포인 트를 기준으로 모델이 어떻게 예측했는지를 설명하며, 특정한 데이터 포인 트에 대한 설명을 제공한다. 이 방법은 모델의 예측을 해석하는 것에 유용 하나, 어떤 feature가 모델 예측에 영향을 끼쳤는지, 그리고 영향력이 얼마 나 큰지에 대한 정보를 제공하지 않는다는 한계가 있었다 [13].

Integrated Gradients는 모델의 입력과 출력 간의 미분을 사용하여 모델 예측의 각 feature에 대한 기여도를 계산한다. 그러나 이 방법은 수치적 불 안정성으로 인해 모델 해석에 대한 일관성이 떨어지며, 시간이 많이 소요 되기 때문에 더 복잡한 모델에서는 계산이 어려울 수 있다 [14].

Partial dependence plot (PDP)은 머신러닝 모델의 각 feature가 모델의 예

측에 미치는 영향을 시각하는 기법으로 PDP plot은 하나의 feature가 다른 feature와 상호작용하지 않는 가정하에 해당 feature의 값 범위를 변화시켰 을 때 모델 예측값이 어떻게 변화하는지 볼 수 있다. 이를 통해 모델 예측 에 영향을 미치는 feature를 파악할 수 있으나 각 feature간의 상호작용을 고려하지 않아 비교하려는 feature가 많아지면 시각화의 어려움이 있다 [15].

SHAP (SHapley Additive exPlanations)은 XAI (Explainable AI) 분야에서 사용되는 모델 설명 기법이다. Shapley value는 합리적인 분배 방식을 제공 하는 게임 이론에서 유래한 개념으로, 변수들의 중요도를 평가할 수 있다. SHAP은 각 변수가 예측 결과에 얼마나 기여하는지를 계산하여 음의 관 계, 양의 관계까지 영향력의 방향을 수치적으로 설명할 수 있다. 즉, SHAP 을 사용하여 모델을 분석하면 각각의 입력변수의 값이 변할 때 예측값이 어떻게 변화하는지를 해석할 수 있다. SHAP은 LIME과 마찬가지로 개별 데 이터 포인트에 대한 설명을 제공하지만, 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 장 점이 있다 [16].

$$\boldsymbol{\varPhi}_{i} = \sum_{S \subseteq F \setminus i} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup i}(x_{S \cup i}) - f_{S}(x_{S})]$$
(3.2)

Ø<sub>i</sub> : i 데이터에 대한 Shapley Value
F : 전체 집합
S : 전체 집합에서 i번째 데이터가 제외한 모든 부분 집합
f<sub>S∪i</sub>(x<sub>S∪i</sub>) : i번째 데이터를 포함한 전체 기여도
f<sub>s</sub>(x<sub>s</sub>) : i번째 데이터 제외된 나머지 부분 집합의 기여도

이 수식은 Shapley value 공식에서 feature importance 계산을 위해 변형 된 것으로, SHAP value는 해당 feature가 모델 예측값에 기여하는 정도를 나타내는 값을 의미한다. 이 공식에서는 해당 feature가 모델 예측에 어떻 게 기여하는지를 feature subset마다 계산하고, 이를 모두 더해서 최종 SHAP value를 구한다 [16].

## 3.5 XGboost 적용 선박 연료소모량 예측 모델

XGBoost는 Gradient Boosting의 개념을 확장하고 개선한 머신러닝 알고리 즘이다. 기존의 Gradient Boosting에는 느린 속도, 과적합 위험 등의 한계가 있었는데, XGBoost는 이러한 한계를 극복하고 높은 예측 성능을 보인다. XGBoost는 트리 기반 알고리즘을 사용하며, 부스팅 과정에서 가중치가 큰 예측 오차에 더 집중함으로써 모델 성능을 향상시킨다. 또한, 병렬 처리 와 고차원 데이터 처리 기능을 강화하여 대규모 데이터셋에서도 높은 성능 을 보인다. XGBoost는 결측값을 자체적으로 처리하고, 과적합 문제를 해결 하기 위해 regularization을 제공한다. 또한 다양한 하이퍼파라미터를 제공 하여 사용자가 모델을 더욱 세밀하게 제어할 수 있다는 장점이 있다 [17].

#### 3.6 XGboost 적용 선박 연료소모량 예측 모델 평가

본 연구에서는 학습데이터와 평가데이터를 8:2의 비율로 사용하였으며 예 측 모델의 성능 평가 지표로 평균제곱근오차 (Root Mean Squared Error, RMSE), 평균절대오차(Mean absolute error, MAE), *R*<sup>2</sup>(R-squared), 평균백분 율오차 (Mean Absolute Percentage Error, MAPE), 를 사용하여 모델의 성 능을 평가하였다. 평균제곱근오차 (Root Mean Squared Error, RMSE): 예측값과 실제값의 차 이를 제곱한 후, 평균을 구한 뒤 다시 제곱근을 취한 값이다. 예측값과 실 제값의 차이가 클수록 오차가 커지는 특성을 가지며, 작을수록 모델의 예 측력이 좋다고 볼 수 있다.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (X_i - Y_i)^2}$$
(3.3)

MAE(Mean absolute error): 예측 모델의 성능을 평가하기 위한 지표 중 하나로 실제값과 예측값의 차이를 평균한 값이다. 즉, 값이 작을수록 모델 의 예측이 더욱 정확하다.

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |X_i - Y_i|$$
(3.4)

R<sup>2</sup>(R-squared): 예측값이 실제값과 얼마나 일치하는지를 나타내는 지표로 예측값과 실제값의 차이를 제곱한 후, 전체 데이터의 분산에 대한 예측값 의 분산 비율을 계산한 값이다. 값이 1에 가까울수록 모델의 예측력이 좋 다고 볼 수 있다.

$$R^{2} = 1 - \frac{\sum_{i=1}^{m} (X_{i} - Y_{i})^{2}}{\sum_{i=1}^{m} (Y_{i} - Y_{i})^{2}}$$
(3.5)

평균백분율오차 (Mean Absolute Percentage Error, MAPE): 실제값과 예측

값의 오차를 백분율로 변환하여 평균한 값으로, 예측 대상의 값이 0에 가 까울 때 효과적인 평가 지표이다. 예측값과 실제값의 상대적인 차이를 측 정하며, 작을수록 모델의 예측력이 좋다.

$$MAPE = \frac{1}{m} \sum_{i=1}^{m} |\frac{Y_i - X_i}{Y_i}|$$
(3.6)

본 연구에서는 XGBoost 기반으로 연료소모량 예측 모델을 생성하였으며, GridSearchCV(Grid Search Cross-Validation)를 활용하여 하이퍼파라미터 튜 닝을 진행하였다. 모델의 성능 평가 결과는 과적합(overfitting) 여부를 확인 하기 위해 예측 모델의 성능을 학습데이터와 테스트데이터로 각각 모델을 평가하였으며 결과를 Table 3.7, Table 3.8과 같이 정리하였다.

Table 3.7 The performance evaluation of the model using the training dataset

	RMSE	MAE	R-squared	MAPE
1	77.9982	49.30420	0.9244	0.0250

Table 3.8 The performance evaluation of the model using the test datase

	RMSE	MAE	R-squared	MAPE
1	89.9320	54.9193	0.8989	0.0277

학습용 데이터와 평가용 데이터를 사용해 모델을 평가한 결과 모델의 과적 합 문제가 발생하지 않았고 예측 정확도 89.9320%를 확인하였다. 학습용 데 이터와 평가용 데이터를 사용해 모델을 평가한 결과를 각각 Figrue 3.9, Figure 3.10과 같이 나타내었다.



Figure 3.10 The performance metrics of the model evaluated on the test dataset

#### 3.7 SHapley Additive exPlanation(SHAP)을 통한 모델 해석

목차 3.6에서 생성한 연료소모량 예측 모델을 기반으로 SHapley Additive exPlanation(SHAP)을 활용하여 예측 결과에 대한 변수들의 영향을 분석하 였다. 먼저 SHAP 사용하여 전체 데이터에 대한 식 3.7을 이용하여 전체에 서 특성당 절대 Shaply value 평균을 구한 Global Importance를 분석하였 다. 이를 통해 연료소모량 예측 모델에 사용한 feature가 전체 모델 출력에 어떻게 영향을 미치는지를 분석하였다.



Figure 3.11 mean absolute SHAP value

Figure 3.11에서 같이 연료소모량 예측 모델에 영향을 미친 feature들의

SHAP value 절대값 평균을 확인할 수 있으며 선박의 대지속도(Speed of Groud; SPEED\_VG), 선박의 상대풍속(Relative wind speed; REL\_WIND\_SPEED), 선박의 평균흘수(Average Draft;AvgDraft), Trim, Heel, 너울의 높이(SWELL\_WAVE\_HEIGHT) 순서로 비교적 높은 영향도를 확인하였다. Figure 3.12에 제시된 SHAP value plot은 각 feature의 SHAP Value와 해당 값이 모델의 예측 결과에 미치는 영향을 나타내는 색으로 구분하여 보여준다. 붉은색은 feature들의 값이 높을수록 모델의 예측 결과에 큰 영향을 미치며, 파란색은 feature들의 값이 낮을수록 모델의 예측 결과에 큰 영향을 미치다는 것을 의미한다. 연료소모량 예측 모델에 각 feature들이 미치는 양의관계, 음의관계를 Table 3.9과 같이 정리하였다.



Figure 3.12 SHAP values(impact on model output)

Table 3.9Summarizing the relationship between the prediction model and<br/>features through the SHAP value plot.

	Relationship	Feature name
1	Positive(red)	SPEED_VG, REL_WIND_SPEED, SWELL_WAVE_HEIGHT, WIND_WAVE_HEIGHT, WIND_WAVE_PERIOD
2	Negative(blue)	Trim, Heel

 Figure 3.11, Figure 3.12를 종합적으로 정리하여 분석하면 선박의 속도가

 클수록, 상대 풍속의 클수록, 파고의 높이가 높을수록, 너울의 높이가 높을

 수록 선박 연료소모량이 증가할 가능성이 있다. 선박의 Trim과 Heel이 높

 을수록 모델의 결과에 부정적 영향을 미쳤다. 즉 선박의 기울어짐이 클수

 록 연료소모량이 증가할 가능성이 있다. Figure 3.11에서 선박의 연료소모

 량 예측에 절대적인 영향을 미친 요인 3가지에 대하여 feature간의

 Dependence plot 분석을 Figure 3.13-20과 같이 진행하였다. SPEED\_VG는

 REL\_WIND\_SPEED, SWELL\_WAVE\_PERIOD, SWELL\_WAVE\_HEIGHT와 높

 은 상호의존성을 가지며, REL\_WIND\_SPEED는 AvgDraft, SPEED\_VG, Trim

 과 높은 상호의존성을 가지며, AvgDraft는 Trim, WIND\_WAVE\_PERIOD,

 SPEED\_VG와 높은 상호의존성을 가진다.









Figure 3.20 Dependence plot of AvgDraft and SPEED\_VG

또한 3가지의 feature와 선박의 연료소모량의 상호 의존도와 부분의존도 를 Figure 3.21-3.26과 같이 나타낼 수 있다. 즉, 종합하면 있어서 선박의 속도가 13knot 이상일 경우, 상대풍속이 11knot 이상, 선박 평균 흘수가 50m 이상일 때 연료소모량 예측값 증가에 기여함을 알 수 있다.



Figure 3.21 Dependence plot of SPEED\_VG

Figure 3.22 PDP plot of SPEED\_VG



Figure 3.25 Dependence plot of Figure 3.26 PDP plot of AvgDraft AvgDraft

하지만 Figure 3.21-26의 결과로 선박 연료소모량의 증감에 대해서는 알 수없다. 따라서, 본 연구에서는 연료소모량이 크게 발생한 관측데이터와 전 체 데이터간의 SHAP Value 비교를 진행하였다. 본 연구에서는 Figure 3.27 과 같이 ME1\_FOC 분포를 통해 ME1\_FOC의 Density는 낮고 ME1\_FOC는 높은 포인트인 ME1\_FOC가 2750이상인 구간의 데이터를 분석하여 ME1\_FOC가 높게 나타는 원인을 분석하였다.



모든 데이터의 plot을 Figure 3.28과 같이 나타냈으며 파란색은 해당 구간 의 FOC 값을 나타내고, 빨간색은 threshold 값인 2750 이상인 FOC값을 나 타낸다. ME1\_FOC가 2750이상인 붉은색 구간에 대해서 전체 데이터와의 SHAP Value 비교 분석을 진행하였다.



ME1\_FOC가 2750이상인 관측치는 152개로 나타났으며, Figure 3.29의 붉 은색 표기와 같이 남아프리카의 케이프타운을 근처로 밀집한 것으로 확인 하였다.



Figure 3.29 Map showing observations with high foc (red dots)

ME1\_FOC가 2750이상인 관측치에 대해서 Figure 3.30과 같이 Force Plot

을 통해 각각의 관측치별로 예측값 기여도를 나타내었다. x축은 각 데이터 관측치이며 y축은 연료소모량이다. 그래프를 통해 AvgDraft는 연료소모량 을 낮추는 REL\_WIND\_SPEED, SPEED\_VG, WIND\_WAVE\_HEIGHT, WIND\_WAVE\_PERIOD는 연료소모량을 증가시키는 요소들로 발견됨음 확인 할 수 있다.



Figure 3.30 Force Plot with ME1\_FOC greater than 2750

SHAP value의 평균을 계산하여 남아프리카 지역과 전체 데이터셋의 SHAP value를 비교하기 위하여 이를 데이터프레임 형태로 만들어 Table 3.10과 같이 SHAP 평균값을 비교하였다. 전체 데이터셋에서는 SPEED\_VG, Trim가 높을수록 연료소모량 증가하였고 REL\_WIND\_SPEED가 낮을수록 연 료소모량이 증가하였다. 하지만, 연료소모량이 크게 발생한 남아프리카 지 역 데이터셋에서는 REL\_WIND\_SPEED, SPEED\_VG, WIND\_WAVE\_HEIGHT, WIND\_WAVE\_PERIOD, SWELL\_WAVE\_PERIOD, CURRENT\_SPEED 의 SHAP Value가 특히 크게 나타났으며 이를 통해 해당 feature들의 값이 컸 고 이로 인해 연료소모량이 크게 발생하였음을 알 수 있었다. 특히, 선박의 외부 조건인 WIND\_WAVE\_HEIGHT, WIND\_WAVE\_PERIOD, SWELL\_WAVE\_PERIOD, CURRENT\_SPEED가 전체 데이터셋에 비해서 크게 기여했음을 알 수 있다.

	features	South_africa	All
0	SPEED_VG	190.759094	14.258555
3	Trim	24.453243	8.568977
8	WIND_WAVE_PERIOD	96.488419	1.981623
13	SWELL_WAVE_PERIOD	70.867287	1.318987
9	CURRENT_SPEED	27.607906	0.989486
5	RUDDER_ANGLE	17.998415	0.696779
11	SWELL_WAVE_HEIGHT	-9.667458	0.607939
6	WIND_WAVE_HEIGHT	121.321915	0.232600
10	TIDEWAY	0.234860	-0.023434
2	LEEWAY	11.823912	-0.035883
15	REL_WIND_ANGLE	-0.091090	-0.134646
7	WIND_WAVE_ANGLE	7.019795	-0.264497
12	SWELL_WAVE_ANGLE	1.337704	-0.662614
4	Heel	11.916217	-2.658451
14	REL_WIND_SPEED	236.029953	-10.235121
1	AvgDraft	-39.065880	-1

Table 3.10 Comparison of shap value

남아프리카 인근 관측치들은 전체 데이터셋에 비해 FOC 예측에 있어서 SPEED\_VG, Trim, WIND\_WAVE\_PERIOD, SWELL\_WAVE\_PERIOD, CURRENT\_SPEED, RUDDER\_ANGLE, SWELL\_WAVE\_HEIGHT, WIND\_WAVE\_HEIGHT, LEEWAY, WIND\_WAVE\_ANGLE, SWELL\_WAVE\_ANGLE, Heel, REL\_WIND\_SPEED가 더 큰 영향 을 끼쳤음을 알 수 있으며 이를 Figure 3.31의 South africa(파란색) 그래프를 통해 확 인할 수 있다.



ME1\_FOC 가 가장 높았던 개별 포인트(Index:15844)의 SHAP force plot을 통 해 feature들의 기여도를 Figure 3.32-3.33과 같이 나타냈다. REL\_WIND\_SPEED, WIND\_WAVE\_HEIGHT, SPEED\_VG, WIND\_WAVE\_PERIOD, SWELL\_WAVE\_PERIOD, Trim, RUDDER ANGLE, CURRENT\_SPEED가 ME1\_FOC를 높이는데 기여하였다.



Figure 3.32 SHAP Force Plot



Figure 3.33 SHAP Waterfall Plot

모델의 해석 결과의 타당성을 더욱 높이기 위하여 남아프리카 지역의 feature별 분포 그래프를 Figure 3.34, Figure 3.35 와 같이 나타내었다. 해당 구간에서 ME1\_RPM은 비교적 58~58.5사이로 일정한 것으로 보아 해당 구간은 Feature들의 영향으로 ME1\_FOC가 높았던 것으로 생각할 수 있다.

각 Feature들의 분포를 살펴보면 IND\_WAVE\_PERIOD, SWELL\_WAVE\_PERIOD 가 증가함을 확인할 수 있다. 파도와 너울의 주기가 통상적으로 선박 운항에 큰 영향을 미치지는 않지만 보통 파도나 바람이 더 멀리서 형성되어 강력한 바람이 화경적인 조건이 않았음을 추측해볼 수 불며 좋지 있다. 실제로 WIND\_WAVE\_HEIGHT이 높아지고 곧 이어 SWELL\_WAVE\_HEIGHT도 높아져 선 박 운항에 안좋은 영향을 미쳤을 것으로 보인다. 또한 WIND\_WAVE\_ANGLE과 SWELL\_WAVE\_ANGLE의 분포를 살펴보면 해당 구간에서 방향이 급변함에 따라 RUDDER\_ANGLE이 영향을 받아 변화하는 것을 확인할 수 있다. 이로 인하여 선박 은 운항에 안좋은 영향을 받았을 것으로 보이며 이는 SPEED\_VG가 해당 구간에 떨 어진 것을 통해 해당 구간의 바람, 너울의 영향이 컸다는 커서 선박 연료소모량이 증가하였다고 생각할 수 있다.

또한 TIDEWAY가 순간적으로 튀는 구간에서 ME1\_LOAD도 감소하며

CURRENT\_SPEED 도 감소함을 확인할 수 있다. 이를 통해 바람, 너울 뿐만 아니라 조류의 흐름이 역조였을 것으로 추측해볼 수 있으며, 이러한 요인이 선박 연료소모 량을 증가시켰을거라고 추측해볼 수 있다.



Figure 3.34 Plots each feature in South Africa section (1)



Ⅳ. 결론

본 연구에서는 16만톤급의 벌크선 데이터를 사용하여 선박 메인엔진의 연 료소모량 예측 모델을 개발하고 SHAP(SHapley Additive exPlanations)을 통 하여 예측 모델의 결과를 해석하였다. 모델의 해석과 설명력을 높이기 위 해 새로운 변수 변환 과정을 거쳤으며 전체 데이터셋에 대한 Global importance와 ME1\_FOC가 가장 크게 나타는 지역 데이터셋에 대한 SHAP 결과를 비교하였다.

모델의 예측 정확도는 RMSE 값은 89.9320, MAE 값은 54.9193, R-squared 값은 0.8989, MAPE 값은 0.0277의 모델 정확도를 확인할 수 있었다.

본 연구에서는 선박의 속도, 상대 풍속, 파고의 높이, 너울의 높이와 같은 요인이 선박의 연료소모량에 영향을 미친다는 것을 발견했다. 또한 선박의 Trim과 Heel이 높을수록 연료소모량이 증가할 가능성이 있다는 것을 확인 했다. 또한, Feature간의 상호의존성을 분석한 결과, SPEED\_VG, REL\_WIND\_SPEED, SWELL\_WAVE\_PERIOD, SWELL\_WAVE\_HEIGHT, AvgDraft, Trim, WIND\_WAVE\_PERIOD와 같은 요인들이 연료소모량에 큰 영향을 미치는 것으로 나타났다. 따라서 선박의 속도가 13knot 이상일 경우, 상대풍속이 11knot 이상, 선박 평균 흘수가 50m 이상일 때 연료소모량 예측값 증가에 기여 함을 알 수 있다.

또한, 본 연구에서는 선박 연료소모량이 가장 크게 나타난 남아프리카 데이터 셋과 전체 데이터셋과의 차이를 비교하였다. 남아프리카 지역에서는 특히 외부 환경 조건에 대한 feature들이 더 큰 영향을 미쳤다는 것을 확인할 수 있었다. 또한, ME1\_FOC가 가장 높았던 개별 포인트에서는 REL\_WIND\_SPEED, WIND\_WAVE\_HEIGHT, SPEED\_VG 등이 FOC를 높이는데 기여하였으며, 해당 구간에서 파도와 너울의 주기가 더 컸고 RUDDER\_ANGLE도 크게 변화하였다는

- 53 -

것을 분석 결과를 통해 확인할 수 있었다. 마지막으로, 각 feature별 분포를 살 펴보면 해당 구간에서의 환경적인 조건의 악화와 선박의 조종 어려움 등이 연 료소모량의 증가에 큰 영향을 미쳤다는 것을 추측할 수 있었다. 이러한 결과는 선박 운항 관리에 있어서 매우 유용하며, 선박 운항을 보다 효율적으로 관리하 여 연료비를 절감하고, 친환경적인 선박 운항을 실현할 수 있는 데에 기여할 수 있으며 이번 분석에서 사용한 모델과 해석 방법론은 다음과 같은 유사한 문 제에 적용될 수 있을 것으로 기대된다.

- 1) 연료소모량에 큰 영향을 미치는 요인 중 AvgDraft의 경우에는 화물의 적재 상태에 따라 달라지는 지표로 향후 항로별 운항 기준을 수집하여 연료소모 량에 대한 기여도에 관한 연구
- CI 등급의 효율적 달성 및 유지를 위한 실시간 데이터 기반의 최적의 속도 예측에 관한 연구
- 3) 다양한 분석 모델 기반의 연료소모량 예측과 해상 시운전 데이터 기반 예 측 모델과의 비교 분석을 통한 예측 모델의 feature 기여도 분석 정확도 향 상 연구
- 4) CII 등급을 효율적으로 달성하기 위한 실시간 연료소모량 예측과 기여 feature 분석과 CII등급간의 오차범위에 관한 연구

# 참고문헌

- [1] 국제에너지기구(IEA). (2021). Monthly Oil Market Report. [Online]. Available:https://www.iea.org/reports/monthly-oil-market-report. [Accessed: 02-May-2023].
- [2] International Maritime Organization. (2020). Fourth IMO GHG Study 2020: Full Report. London: International Maritime Organization.
- [3] 한국선급(2021). CII 규제 대응 지침서. 한국선급.
- [4] 현대해양, 선박연료 교체 위한 정부 정책 방향, http://www.hdhy.co.kr/news/articleView.html?idxno=17162
- [5] Cheon, J. (2019). A study on the evaluation of Greenhouse gas from ship's alternative fuel considering the life cycle assessment. (Doctoral dissertation, Pukyong National University, Department of Mechanical System).
- [6] olland, A. F., Turnock, S. R., Hudson, D. A., & van Gelder, P. A. C. M. (2011). Ship resistance and propulsion. Cambridge University Press.
- [7] MTaylor, D. A. (1980). Marine Engineering: Thermodynamics and Heat Engines. Prentice-Hall.
- [8] Kim, Y., & Lee, K. (2019). Machine learning approach to ship fuel consumption: A case of container vessel. Journal of Marine Science and Engineering, 7(6), 179.
- [9] Kim, D., Handayani, M. P., Lee, S., & Lee, J. (2021). Feature Attribution Analysis to Quantify the Impact of Oceanographic and Maneuverability Factors on Vessel Shaft Power Using Explainable Tree-Based Model. Journal of Marine Science and Engineering, 9(5), 530.

- [10] Lipton, Z. C. (2016). The mythos of model interpretability. arXiv preprint arXiv:1606.03490.
- [11] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.
- [12] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., & Legg, P. A.
   (2019). A Survey of Explainable Artificial Intelligence (XAI). IEEE Access, 7, 14230-14245. doi: 10.1109/ACCESS.2019.2893077
- [13] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should i trust you?: Explaining the predictions of any classifier." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.
- [14] Sundararajan, Mukund, Ankur Taly, and Qiqi Yan. "Axiomatic attribution for deep networks." Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017.
- [15] Hastie, T. J., & Tibshirani, R. J. (1990). Generalized Additive Models. Chapman and Hall.
- [16] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in neural information processing systems (pp. 4765–4774).
- [17] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), pages 785-794, 2016.
- [18] 윤여정. (2015). 지문항해학. 한국해양과학기술원.

# 감사의 글

지난 2년 반 동안 연구원과 학위를 병행하며 이렇게 석사 과정을 마무리하며 감사의 글을 쓰니 감회가 새롭습니다. 석사 과정을 돌아보면 주변에 도움을 주 신 감사한 분들이 정말 많은 것 같습니다.

먼저 저의 논문이 완료되도록 지도해주신 이지환 교수님, 서원철 교수님, 최성 철 교수님께 진심으로 감사의 말씀을 드립니다. 학부 전공과 석사 전공이 달라 져서 부족함이 많은 저를 지도해주시고 연구 주제와 논문 작성 과정에서 하나 하나 세심하게 지도해주시고 아낌없이 가르쳐주신 이지환 교수님께 추가로 감 사의 말씀을 드립니다.

또한, 학위 과정을 추천해주신 책임님께도 진심으로 감사의 말씀을 드립니다. 처음 사회생활을 하면서 지금까지 많은 배려를 해주신 덕분에 업무와 병행하면 서 학위를 잘 마칠 수 있었고, 학위, 연구 등 다방면으로 함께 고민해주시고 조언 과 격려를 해주신 덕에 많은 것을 배우며 성장할 수 있었습니다.

그리고 석사 학위 중에 연구에 많은 도움을 주었던 BA lab 동료들에게도 감사 의 말씀을 전합니다. 자주 만나지는 못했지만 항상 도움을 주는 좋은 동료들을 만나게 되어 감사합니다.

마지막으로 저를 항상 믿고 지지해주는 사랑하는 가족과 학업과 일을 병행하 면서 소홀해도 항상 이해해주고 배려해주는 곧 남편이 될 남자친구에게도 감사 를 전하고 싶습니다.

- 57 -