

# XAI를 활용한 특별관리해역의 연안오염총량관리제 오염물질 대상항목 선정 및 해양오염 관리 방안 도출

김세상<sup>†</sup>

부경대학교 기술데이터공학과 대학생

## Using XAI to Select Pollutants of Total Pollutant Load management System (TPLMS) in Special Management coastal Zone (SMCZ) and Derive Marine Pollution Control Measures

Sesang Kim<sup>†</sup>*Undergraduate Student, Department of Technology and Data Engineering, Pukyong University, Busan 48513, Korea*

### 요 약

해양수산부는 매년 특별관리해역 현황을 보고하고 있다. 현재 우리나라는 주기적으로 해양환경측정망 자료를 활용해 해양오염의 정도를 파악하여 특별관리해역 연안오염총량관리 계획을 수립하고 평가하고 있다. 하지만 해양수산부에서 2016년 해양환경측정망 운영 결과, 대부분 해역이 WQI(Water Quality Index, 수질 평가 지수) 2등급 이상의 비교적 깨끗한 수질을 유지하고 있으나, 특별관리해역 중 울산 연안, 마산만 내측, 시화호 내측, 낙동강 하구 해역과 영산강 하구의 목포 연안 등 육상 오염물질 유입의 직접적인 영향을 받는 일부 정점 등에서는 부분적으로 WQI 4등급 이하의 결과를 보여 수질개선을 위한 대책이 필요한 것으로 나타났다. 이 연구는 해양환경측정망의 특별관리해역 데이터를 활용해 XAI(eXplainable Artificial Intelligence, 설명할 수 있는 인공지능) 분석을 통한 특별관리해역의 해양오염원의 현황을 파악하고, 주 오염원을 도출하여 연안오염총량관리제 오염물질 대상 항목을 선정하는 의사결정 문제에 새로운 방향성을 제안한다. 또한 오염물질별 영향도를 도출해 해양오염지수를 생성하여 향후 특별관리해역 후보지 선정과 같은 정책 수립 방안을 모색하고 해양오염을 관리하기 위해 미래지향적 해결방안을 도출해 볼 수 있었다. WQI, COD(Chemical Oxygen Demand, 화학적 산소 요구량), P(Phosphorus, 인) 등 해양환경에 영향을 미치는 해양환경 데이터를 시계열별로 수집하고 정제하여 현재 특별관리해역에서 해양오염원의 영향도를 XAI를 통해 분석하여 각 해역별 연안오염총량관리제 오염물질 대상 항목 선정 의사결정 문제에 새로운 방향성을 제안하였다. 또한 해양오염원의 영향도를 반영한 해양오염지수들도대로 1년 후 해양오염을 예측하는 AI 모델링을 하여 향후 특별관리해역 후보지를 선정하고 선정된 특별관리해역 후보지 구역을 관리하기 위한 해결방안을 도출해볼 수 있었다.

**Abstract** – The Ministry of Oceans and Fisheries reports on the status of special management coastal zone every year. Currently, Korea periodically utilizes the data from the marine environmental measurement network to identify the extent of marine pollution to establish and evaluate the total pollutant load management system for special management coastal zone. However, as a result of the operation of the marine environmental measurement network in 2016 by the Ministry of Oceans and Fisheries, most of the waters maintain relatively clean water quality with a water quality index (WQI) of 2 or higher. However, among the special management coastal zone, some peaks that are directly affected by the influx of land-based pollutants, such as the coast of Ulsan, the inner part of Masan Bay, the inner part of Sihwa Lake, the waters of the Nakdong River estuary, and the coast of Mokpo at the mouth of Yeongsan River, partially show results below WQI 4, indicating that measures to improve water quality are needed. This study identifies the current status of marine pollution sources in special management coastal zone through XAI (explainable artificial intelligence) analysis using data on special management coastal zone from the Marine Environmental Measurement Network, identifies the main pollution sources, and proposes a new direction

<sup>†</sup>Corresponding author: sstptkdss1@icloud.com



for the decision-making problem of selecting pollutant targets for the total pollutant load management system. In addition, by deriving the impact of each pollutant and creating the marine pollution index, I was able to explore ways to establish policies such as selecting candidate sites for special management coastal zone in the future and derives future-oriented solutions to manage marine pollution. I was able to collect and refine marine environmental data affecting the marine environment, such as WQI, COD (chemical oxygen demand), and P (phosphorus), by time series, and analyze which marine pollutants have the most influence in the current special management coastal zone through XAI to propose a new direction for the decision making problem of selecting pollutant targets for the total pollutant load management system for each area. Also, Based on the marine pollution index, which reflects the impact of marine pollution sources, AI modeling to predict marine pollution one year from now was used to select candidate areas for future special management coastal zone and derive solutions for managing the selected special management coastal zone.

**Keywords:** Machine learning(머신 러닝), Time series(시계열), XAI(설명가능한 인공지능), Marine pollution(해양오염), Marine pollution index(해양오염지수), Total coastal pollutant load management system(연안오염총량관리제), Special management coastal(특별관리해역)

## 1. 서 론

특별관리해역은 「해양환경관리법」 제 15조(환경관리해역의 지정·관리)에 따라 해양 환경 기준의 유지가 곤란한 해역 또는 해양 환경 및 생태계의 보전에 현저한 장애가 있거나 장애가 발생할 우려가 있는 해역으로서 대통령령이 정하는 해역(해양오염에 직접 영향을 미치는 육지를 포함)이다. 따라서 인근에 대규모 산업단지나 대도시가 위치하고 있어 다른 해역에 비해 국민의 건강 및 해양환경의 보전을 위해 특별히 관리가 필요한 해역이다. 이 해역은 오염물질의 총량규제 '특별관리해역 연안오염총량관리'를 실시하고 있다. 이는 해양 환경에서 발생하는 오염 물질의 배출량을 관리하고, 오염물질에 노출되는 인구나 수생생물에 대한 위해성을 최소화하기 위한 다양한 정책과 기술적 조치를 포함한 총괄 관리 체계이다.

「특별관리해역 연안오염총량관리 기본방침」 제 4조(연안오염총량관리 항목의 선정)에서 해양수산부장관은 영 제 12조 제 1항에 따라 총량관리계획기간에 관리해역의 수질에 대한 영향, 법적 규제와의 연계성, 지표로서의 대표성, 감시측정의 용이성 등을 종합적으로 고려하여 연안오염총량관리를 실시하는 해역의 관할 광역시장 및 도지사와의 협의를 거쳐 연안오염총량관리 항목을 결정한다.

이처럼 지금까지 과거의 데이터만으로 수질관리계획을 세웠기 때문에 해당 관리계획이 수질 개선에 미칠 영향을 파악하기 어려웠을 것이다. 이를 뒷받침할 근거로, 환경관리해역에서 국가해양환경정보통합시스템(MEIS) 자료를 활용하여 2006~2017년까지 경년별 수질변화경향을 비교한 결과, 특별관리해역 광양만과 마산만 및 환경보전해역인 가막만에서는 주로 하계에 빈산소수괴(hypoxia water mass, 용존산소의 농도가 낮은 물 덩어리)를 빈산소수괴가 출현했다.

또한 수질 오염도를 나타내는 COD 경우에도 Fig. 1에서 나타나는 것처럼 해양환경측정망 2015~2020년도 데이터에서 특별관리해역 중 광양만과 인천 연안의 연도별 COD는 단기적으로 개선되는 것으로 보였으나, 장기적으로 보았을 때 COD가 꾸준히 증가하는 추세를 보인다. 따라서 당장 직면한 상황을 수습하려는 정책들로 미래의 해양오염을 감소시키는 데에는 한계가 있었을 것이라 예상할 수 있었고, 이는 특별관리해역의 관리목적과 부적합하다.

또한 현재 연안오염총량관리제 시행에 따르면 해양수산부는 상대적으로 오염이 심한 해역의 수질 개선을 위해 해역별로 오염원인 분석 및 추가 정밀조사를 하고, 필요한 경우 특별관리해역으로 추가로 지정하여 연안오염총량관리제를 시행하는 등 집중적으로

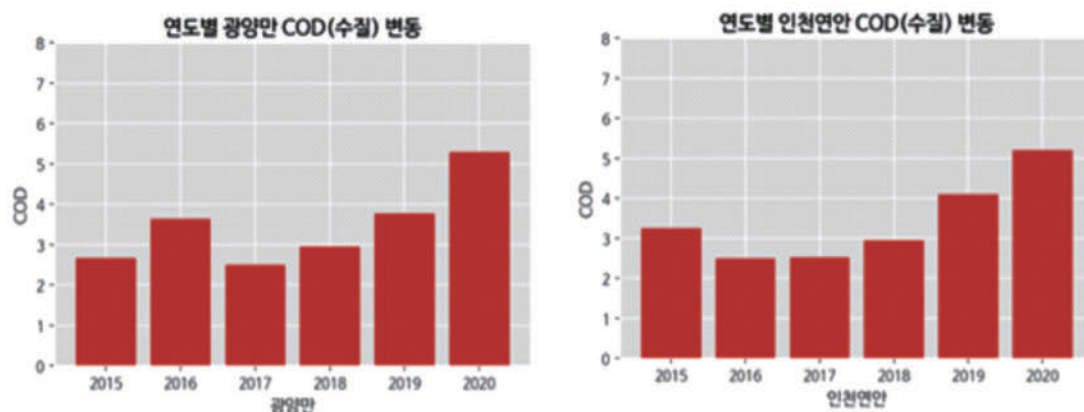


Fig. 1. Gwangyang Bay/Incheon Coast COD graph by year.



**Table 1.** Types of pollutants to be managed by total quantity control plan period by special management coastal zones

연안오염총량관리 시행 특별관리해역	총량관리계획기간	관리대상 오염 물질
마산만	제3차 총량관리계획기간(2017~2021)	화학적산소요구량, 총인
시화호	제2차 총량관리계획기간(2018~2022)	화학적산소요구량, 총인
부산연안	제1차 총량관리계획기간(2015~2019)	화학적산소요구량
울산연안	제1차 총량관리계획기간(2018~2022)	중금속(구리, 아연, 수은)

관리하는 방안을 검토하고 있다. 이러한 관리해역에서의 이용·개발행위가 지속해서 이루어지고 있으므로, 이에 따른 해양환경 부문 환경영향평가에서 충분한 수질 변화에 대한 실태분석 및 사업 추진에 따른 수질 영향과 퇴적물의 오염상태를 철저히 진단해서 오염원 관리에 대한 대책이 중점적으로 평가되어야 할 것이다.

따라서 이러한 연안오염총량관리제 오염물질 선정의 의사결정에 해양환경측정망 과거 데이터 자체만을 활용하여 정책을 만드는 것이 아닌 XAI를 통해 현재 특별관리해역으로 지정된 구역을 분석하여 해양오염 요인들이 얼마나 영향을 미쳤는지 분석하고 해양오염을 예측하는 방법을 활용함으로써 해양오염 관리에 새로운 방향성을 제안하고자 한다. 미래지향적이고 장기적인 효과를 추구하는 정책을 수립하여 더 효과적인 수질 관리 방안을 찾아낼 것이다.

이 연구에서는 2018~2022년도까지 5년의 해양환경 데이터를 활용하여 해안의 해양환경측정망 시계열 데이터 수집한다. 현재 특별관리해역(대분류 6개, 소분류 89개)의 WQI, COD, P 등 다양한 해양오염원 데이터를 수집해 해양오염원에 따른 특별관리해역을 ML 기법을 활용해 Random Forest Classifier 모델로 이진 분류 학습시킨다(Features : 관측 년, 월, 날씨 및 WQI, COD, P 등 32가지의 해양오염원, Target : 특별관리해역 여부(0/1 binary)).

그 후 XAI SHAP를 활용하여 각 특별관리해역(광양만, 마산만, 부산연안, '시화호연안', '인천연안', '울산연안')을 분석해 어떤 해양오염원이 가장 큰 영향을 끼쳤는지 도출하여 해양환경측정망 해양오염에 영향을 끼치는 요인들은 COD, P, WQI 뿐만 아니라 다양한 요인들이 존재하는 것을 확인해보고 연안오염총량관리제 오염물질 대상 항목의 선정 의사결정 문제에 새로운 방향성을 제안한다.

그리고 현재 특별관리해역의 각 세부 해역마다 선별된 해양오염 관리 대처방안 및 해결방안 도출하고 해양오염원의 중요도를 활용

해 가장 중요도가 높은 변수를 1년 후(2023년) 해양오염지수 AI 모델링 예측 및 시각화한다(Features : 관측 년, 월, 날씨 및 WQI, COD, P 등 32가지의 해양오염원, Target : 가장 영향이 높은 해양오염원). 실제 2022년 해양오염원과 예측한 2023년 해양오염원의 증감 비율 확인하고 2022~2023년까지 해양오염지수 증감비율이 가장 높은 상위 10개의 연안 선정 및 시각화한다. 이렇게 예측된 결과를 이용하여 향후 특별관리해역 후보지 선정 등 정책 수립 방안 모색한다.

## 2. 연구 방법

### 2.1 연구 순서도

#### 2.2 2018-2022 5년 해양환경측정망 데이터 수집

해양환경정보포털에서 해양환경에 영향을 미치는 해안의 해양환경측정망 시계열 데이터 수집했다. 데이터는 7500행, 39열로 구성되어 있으며 정보는 Table 2와 같다.

#### 2.3 데이터 전처리

시계열로 데이터를 알아보기 위해 관측년도-관측월-정점명 순으로 데이터 프레임을 정렬하여 해양오염원 Feature를 선정했다. Target은 특별관리구역과 그렇지 않은 구역으로 Target을 설정하여 'SMA (special management coastal zone(area))' 라는 과생 column을 만들었다.

특별관리해역은 광양만, 마산만, 부산연안, 시화호, 인천연안, 울산연안 이므로 이 해역에 해당하면 '1'으로 해당하지 않으면 '0'으로 설정했다. 이를 이진 분류로 예측하는 모델을 만들어 학습시킬 것이다. 그리고 '날씨' column은 범주형 변수이기 때문에 Label

**Table 2.** Columns and marine observation point names information of data

Columns	Marine observation point
39개	375개
생태구명, 정점명, 관측년도, 관측월, 조사 일자, 날씨, 수심, 투명도(m), 수온(°C)표층, 수온(°C)저층, 염분표층, 염분저층, 수소이온농도표층, 수소이온농도저층, 용존산소량(mg/L)표층, 용존산소량(mg/L)저층, 화학적산소요구량(mg/L)표층, 화학적산소요구량(mg/L)저층, 암모니아성질소(µg/L)표층, 암모니아성질소(µg/L)저층, 아질산성질소(µg/L)표층, 아질산성질소(µg/L)저층, 질산성질소(µg/L)표층, 질산성질소(µg/L)저층, 용존무기질소(µg/L)표층, 용존무기질소(µg/L)저층, 총질소(µg/L)표층, 총질소(µg/L)저층, 용존무기인(µg/L)표층, 용존무기인(µg/L)저층, 총인(µg/L)표층, 총인(µg/L)저층, 규산규소(µg/L)표층, 규산규소(µg/L)저층, 부유물질(mg/L)표층, 부유물질(mg/L)저층, 클로로필A(µg/L)표층, WQI등급	가로림만1~6, 가막만1~5, 감포1~4, 강구1~2, 강릉1~7, 강릉남대천하구1, 거제도남안1~4, 거제도동안1~4, 거진1~2, 고성자란만1~3, 고창1~8, 광양만1~12, 구룡포1~2, 군산1~10, 기장1~4, 낙동1~4, 낙동강하구1~10, 남해1~5, 대산1~4, 대정1~3, 도암만1~6, 동해1~4, 득량만1~5, 마산만1~15, 목포1~8, 무안1~3, 보령1~4, 부산1~17, 사천1~4, 삼척1~4, 삼척오십천하구1, 서귀포1~3, 섬진강하구1~10, 성산포1~2, 속초1~4, 시화호1~10, 신안1~2, 신항~2, 아산1~5, 양양1~4, 양양남대천하구1, 여수1~5, 여자만1~3, 영일만1~12, 온산1~9, 완도1~5, 왕피천하구1, 울산1~12, 월포1~2, 인천1~23, 전주포1~8, 제주1~3, 조천1~3, 주문진1~5, 죽변1~6, 진도1~4, 진주만1~13, 천수만1~7, 축산1~2, 태안1~5, 통영1~4, 통영외안1~6, 표선1~2, 한림1~3, 함평만1~4, 해남1~2, 행암만1~4, 후포1~2

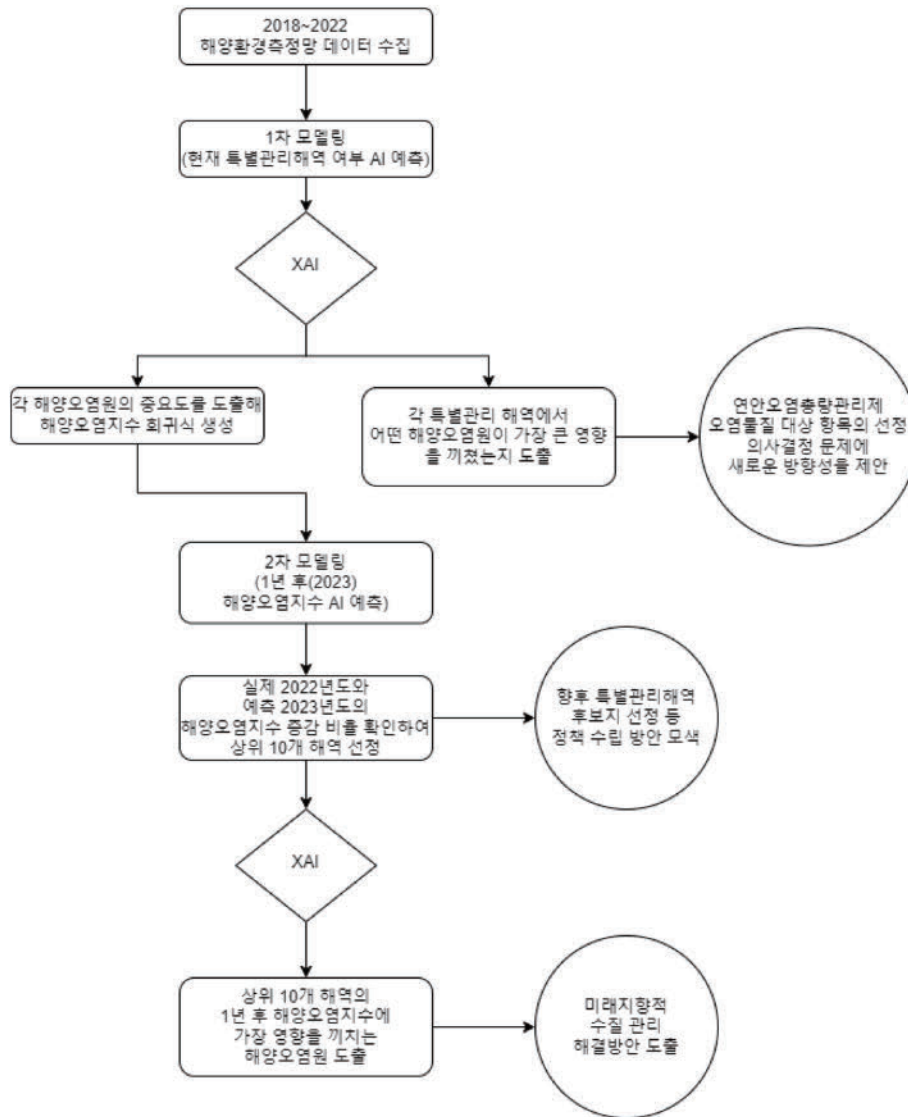


Fig. 2. Research flowchart.

생태구명	점점명	관측도	관측일	조사일자	날씨	수심	투명도 (m)	수온 (°C) 표층	수온 (°C) 저층	...	총인 (µg/L) 표층	총인 (µg/L) 저층	규산규소 (µg/L) 표층	규산규소 (µg/L) 저층	부유물질 (mg/L) 표층	부유물질 (mg/L) 저층	클로로필 A (µg/L) 표층	클로로필 A (µg/L) 저층	WQI 등급	SMA	
0	서해중부	가로림만1	2018	2	2018-02-08	7	14.0	1.6	0.25	0.21	...	29.98	33.66	332.00	328.36	49.80	24.20	0.85	0.96	2	0
1	서해중부	가로림만2	2018	2	2018-02-08	7	24.0	1.5	2.34	2.10	...	41.54	57.67	367.30	367.02	22.30	53.30	0.84	1.0	2	0
2	서해중부	가로림만3	2018	2	2018-02-01	7	25.0	0.9	3.35	3.31	...	38.72	33.20	371.14	361.96	25.40	36.70	0.12	0.44	2	0
3	서해중부	가로림만4	2018	2	2018-02-08	7	17.0	2.5	-0.15	-0.22	...	33.33	32.49	347.70	329.45	9.25	19.76	2.20	1.93	2	0
4	서해중부	가로림만5	2018	2	2018-02-08	7	11.0	1.6	0.74	0.70	...	34.01	33.16	340.90	339.84	18.30	13.10	0.84	0.82	3	0

Fig. 3. Data frame representing the top 5 data.



Table 3. Features and Target in a Binary Classification Model

Feature	Target (SMA)
35개	2개
관측년도, 관측월, 날씨, 수심, 투명도(m), 수온(°C)표층, 수온(°C)저층, 염분표층, 염분저층, 수소이온농도표층, 수소이온농도저층, 용존산소량(mg/L)표층, 용존산소량(mg/L)저층, 화학적산소요구량(mg/L)표층, 화학적산소요구량(mg/L)저층, 암모니아성질소(μg/L)표층, 암모니아성질소(μg/L)저층, 아질산성질소(μg/L)표층, 아질산성질소(μg/L)저층, 질산성질소(μg/L)표층, 질산성질소(μg/L)저층, 용존무기질소(μg/L)표층, 용존무기질소(μg/L)저층, 총질소(μg/L)표층, 총질소(μg/L)저층, 용존무기인(μg/L)표층, 용존무기인(μg/L)저층, 총인(μg/L)표층, 총인(μg/L)저층, 규산규소(μg/L)표층, 규산규소(μg/L)저층, 부유물질(mg/L)표층, 부유물질(mg/L)저층, 클로로필A(μg/L)표층, WQI등급	0, 1

Encoding 기법을 사용해 숫자형 카테고리 값으로 변환했다. 다음 Fig. 3은 상위 5개 데이터를 나타내는 데이터 프레임 예시이다.

2.4 1차 모델링

2.4.1 설명 변수와 목적 변수

이진 분류 모델은 두 개의 클래스 중 하나에 데이터 포인트를 할당하는 ML 모델이다. 일반적으로 이진 분류 모델은 입력 특징 벡터를 입력으로 받아 각 클래스에 대한 확률 값을 출력한다. 확률 값은 모델이 예측한 클래스가 정확한지에 대한 신뢰도를 나타내며, 일반적으로 모델의 출력이 0.5보다 크면 해당 클래스로 예측된다. 이진 분류 모델의 설명 변수(Feature)와 목적 변수(Target)은 Table 3과 같이 설정했다.

2.4.2 모델 선택

이진 분류 모델의 일반적인 알고리즘에는 로지스틱 회귀(Logistic Regression), 랜덤 포레스트(Random Forest) 등이 있다. 모델의 성능을 평가하고 모델링이 타당한지 판별하기 위하여 4가지 ML 모델을 이용해 보았다.

2.4.2.1 로지스틱 회귀(Logistic Regression)

로지스틱 회귀는 선형 회귀와 비슷하지만, 출력이 이진 결과인 점에서 차이가 있다.

로지스틱 회귀 모델은 입력 특징과 가중치 벡터를 곱한 값에 바이어스를 더한 값에 로지스틱 함수를 적용하여 이진 결과를 예측한다. 로지스틱 함수는 S자 형태의 곡선을 그리며, 입력 값이 커질수록 1에 가까워지고 작아질수록 0에 가까워진다.

로지스틱 회귀는 기본적으로 최대 가능도 추정(Maximum Likelihood Estimation)을 사용하여 모델을 학습합니다. 이는 주어진 데이터에서 모델이 예측한 결과와 실제 결과의 차이가 최소화 되도록 가중치를 조정하는 방식이다. 로지스틱 회귀의 우도 함수는 다음과 같이 쓸 수 있다.

$$L = \prod_i \sigma(\beta^T x_i)^{y_i} (1 - \beta^T x_i)^{1 - y_i}$$

2.4.2.2. 랜덤 포레스트(Random Forest)

의사결정트리(Decision Tree)를 기반으로 하는 앙상블 기법 중 하나이다. 여러 개의 의사결정트리를 생성하여 그들의 예측 결과를 평균내어 최종 예측을 수행한다. 랜덤 포레스트는 각 트리를 생성

할 때, 무작위로 선택된 변수들을 이용하여 트리를 학습시키는 방법을 사용한다.

이러한 무작위성(Randomness)은 모델이 과적합(Overfitting)되지 않도록 하며 모델의 일반화 성능을 향상한다. 랜덤 포레스트는 여러 개의 의사결정트리를 앙상블하여 최종 예측을 수행하기 때문에, 단일 의사결정트리보다 높은 예측 성능을 보인다. 또한 랜덤 포레스트는 무작위성을 이용하여 과적합을 방지한다. 각 트리를 학습시킬 때, 무작위로 선택된 변수들만 사용하므로 모델이 특정 변수에 과도하게 의존하는 것을 방지할 수 있다.

2.4.2.3 XGBoost(eXtreme Gradient Boosting)

XGBoost는 그래디언트 부스팅 알고리즘을 사용하는 머신 러닝 라이브러리로 대용량 데이터 셋에 대한 고성능 예측 모델을 생성할 수 있다. XGBoost 모델은 여러 개의 결정 트리를 연속적으로 학습하여 예측 모델을 생성한다.

이러한 결정 트리는 데이터의 특징을 사용하여 이진 분류를 수행하며, 각 노드에서 최적의 분할을 선택하여 분류를 수행한다. 그라디언트 부스팅 알고리즘을 사용하여 각 결정 트리의 가중치를 조정하며, 오분류된 데이터 포인트에 대해 더 강력한 가중치를 부여하여 모델의 성능을 개선하고 대용량 데이터 처리 및 병렬 처리 기능을 제공하여 빠른 모델 학습 및 추론을 지원한다.

2.4.2.4 LGBMClassifier (Light Gradient Boosting Model)

LGBMClassifier는 LightGBM 라이브러리를 사용하는 이진 분류 모델이다. LightGBM은 마이크로소프트에서 개발한 고속 그라디언트 부스팅 프레임워크로 대용량 데이터셋에서 빠른 속도와 높은 정확도를 보장한다.

LGBMClassifier 모델은 기본적으로 그래디언트 부스팅 트리(Gradient Boosting Tree) 알고리즘을 사용하여 작동한다. 이 알고리즘은 여러 개의 결정 트리를 연속적으로 학습하여 모델의 성능을 향상시키고 이러한 결정 트리는 데이터의 특징을 사용하여 이진 분류를 수행하며, 각 노드에서 최적의 분할을 선택하여 분류를 수행한다.

2.4.3 모델 평가

2.4.3.1 Cross val score

Cross\_val\_score는 scikit-learn 라이브러리에서 제공하는 교차 검증 함수 중 하나이다. 이 함수를 사용하면 모델의 성능을 검증하기 위해 데이터 셋을 훈련/검증 세트로 나누는 과정을 반복하여 교차



검증을 수행할 수 있다. cv 매개 변수를 이용하여 몇 개의 fold로 교차 검증을 수행할지 결정할 수 있다. 각 fold에서 모델의 성능을 측정 후, 이를 평균 내어 최종적인 성능 점수를 계산하여 반환한다. 이 점수를 통해 모델의 일반화 성능을 평가할 수 있다. 본 연구의 모델링에서는 cv=5로 설정하면 5-fold 교차 검증을 수행했다.

#### 2.4.3.2 Confusion Matrix

분류 모델의 성능을 평가하는 데 사용되는 행렬이다. 모델의 예측 결과와 실제 결과를 비교하여 모델이 얼마나 정확하게 분류를 수행하는지를 확인할 수 있다. Confusion Matrix는 다음과 같은 네 가지 항목으로 구성된다. True Positive(TP): 실제 값이 Positive이고, 모델이 Positive로 예측한 경우의 수, False Positive(FP): 실제 값이 Negative이지만, 모델이 Positive로 예측한 경우의 수, False Negative(FN): 실제 값이 Positive이지만, 모델이 Negative로 예측한 경우의 수, True Negative(TN): 실제 값이 Negative이고, 모델이 Negative로 예측한 경우의 수.

이러한 Confusion Matrix를 통해 분류 모델의 성능 평가 지표인 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score 등을 계산할 수 있다. 정확도는 전체 데이터 중 모델이 정확하게 분류한 데이터의 비율을 나타내며, 다음과 같은 식으로 계산된다. Accuracy = (TP + TN) / (TP + FP + FN + TN). 또한, 정밀도는 모델이 Positive로 예측한 데이터 중 실제 Positive인 데이터의 비율을 나타내며, 다음과 같은 식으로 계산된다. Precision = TP / (TP + FP). 재현율은 실제 Positive인 데이터 중 모델이 Positive로 예측한 데이터의 비율을 나타내며, 다음과 같은 식으로 계산된다. Recall = TP / (TP + FN). 이러한 성능 지표를 이용하여 모델의 성능을 평가하고, 모델 개선에 활용할 수 있다.

## 2.5 XAI(Explainable Artificial Intelligence)

XAI를 활용하여 특별관리해역 선정에 어떤 해양오염원이 가장 큰 요인인지 도출할 것이다. XAI란, 인공지능 모델의 결과를 해석하고 이를 설명 가능한 형태로 제공하는 기술을 말한다.

보편적으로 인공지능 모델은 많은 양의 데이터를 학습하고, 복잡한 패턴을 인식해 예측을 수행한다. 그러나 이러한 모델은 대개 내부 동작을 설명하기 어렵다. XAI는 이러한 문제점을 해결하기 위해 설명 가능한 인공지능을 개발하는 기술로, 인공지능 모델의 예측 결과를 왜 그런 결과가 나왔는지 설명 가능한 형태로 제공하여, 모델의 신뢰성을 높이고, 모델의 의사 결정을 해석하고 개선할 수 있도록 한다. XAI 기술은 다양한 방법으로 구현된다.

대표적인 방법으로는 SHAP(Shapley Additive exPlanations), LIME(Local Interpretable Model-agnostic Explanations), PDP(Partial Dependence Plot), ICE(Individual Conditional Expectation)가 있다.

이러한 방법들은 모델의 예측 결과를 분석하고, 어떤 변수가 모델의 예측에 영향을 미치는지, 어떤 패턴이 모델의 예측 결과에 기여하는지 등을 분석하여 모델의 예측을 설명 가능한 형태로 제공한다. 본 연구에서는 Random Forest 모델에 SHAP 기법을 이용해 XAI 분석을 수행했다.

### 2.5.1 TreeSHAP

TreeSHAP는 Shapley Additive exPlanations의 한 종류로, 트리 기반 모델에서 각 변수가 모델 예측에 어떻게 영향을 미치는지 설명 가능한 형태로 제공하는 방법 중 하나이다. TreeSHAP[Lundberg et al.(2018)]은 의사결정 트리, 랜덤 포레스트 및 경사 부스트 트리(Gradient Boosted Tree)와 같은 트리 기반 기계학습 모델을 위한 TreeSHAP를 제안했다. TreeSHAP는 트리 모델의 구조와 개별 특성의 값에 대한 정보를 기반으로, 특정 샘플의 예측값을 분해하여 각 변수의 영향력을 계산한다. 트리 모델에서 각 특성이 분기에 영향을 미치는 정도를 계산하고, 모든 가능한 특성 조합에 대한 기여도 값을 계산한다. 이후, Shapley Value로 각 특성이 전체 모델 예측값에 기여하는 정도를 계산한다. Shapley Value  $\phi_i$ 의 계산식은 다음과 같다.

$$\phi_i = \sum_{S \subseteq F \setminus i} \frac{|S|!(|F|-|S|-1)!}{|F|!} [F_{S \cup i}(x_S \cup i) - f_S(x_S)]$$

F: 모든 Features의 집합

f: 예측모델

i = Feature

이렇게 계산된 TreeSHAP 값은 각 변수가 모델 예측에 미치는 영향력을 설명 가능한 형태로 제공한다. TreeSHAP는 조건 기댓값  $E[f(z)|z_s]$ 을 사용하여 효과를 추정한다. Input s와 Feature Set인 S와 한 Tree 모델이 주어졌을 때,  $E[f(z)|z_s]$ 을 계산하는데  $x_i$ 가 속하는 Leaf node의 score값과 해당 node에 속하는 Training 데이터의 비중을 곱하여 계산된다. 그 후 Ensemble 모델 Score를 계산할 때와 마찬가지로, 각 Tree의  $E[f(z)|z_s]$ 을 합하여 Ensemble 모델에 대한 최종  $E[f(z)|z_s]$ 을 계산한다.

### 2.5.2 특별관리해역의 각 해양오염원의 영향도 도출 : Force\_plot

force\_plot은 SHAP에서 제공되는 함수 중 하나이다. 이 함수는 특정 샘플에 대한 SHAP 값을 시각화하여 해당 예측의 설명을 제공한다. 각 피쳐의 SHAP 값은 해당 피쳐가 모델의 예측에 어떤 방식으로 기여하는지를 보여준다. force\_plot은 다음과 같은 작업을 수행한다.

1. 선택한 샘플의 특징(feature) 값들을 기반으로 SHAP 값을 계산.
2. 각 피쳐의 SHAP 값을 시각화하여 해당 피쳐가 어떻게 예측에 영향을 미치는지 시각화.
3. 모델의 바이어스(bias) 향을 포함하여 모델의 예측에 어떤 피쳐가 어떤 정도로 기여도 도출.
4. 모든 SHAP 값의 합이 모델의 예측값과 일치하도록 조정.

이를 통해 force\_plot은 단일 예측의 경우 모델의 동작을 설명하고, 각 피쳐가 예측에 어떻게 기여하는지 이해하는 데 도움을 준다. 이를 이용해 각 컷 번째 특별관리해역에서 해양오염원의 영향도를 도출했다.

### 2.5.3 각 해양오염원의 특성 중요도 도출 : summary\_plot

summary\_plot은 SHAP의 기능 중 하나로, 모델의 전체적인 동



작을 요약하여 시각화하는 데 사용되는 함수이다. 이 함수는 여러 샘플의 SHAP 값들을 기반으로 피처의 중요도를 시각화하여 모델의 예측 설명을 요약한다. summary\_plot에서 제공되는 정보는 다음과 같다.

1. 전체적인 피처 중요도: 모델이 어떤 피처를 주로 활용하는지를 나타낸다.
2. 피처의 영향 방향: 각 피처가 어떤 방향으로 예측에 영향을 미치는지를 나타낸다. 어떤 피처가 양의 영향을 주는지, 음의 영향을 주는지 등을 확인할 수 있다.
3. 영향의 크기: 각 피처의 영향력이 얼마나 강한지를 나타내고 피처의 SHAP 값의 크기로 나타낸다.

2.5.4 MPI(Marine Pollution Index, 해양오염지수) 회귀식 생성 summary\_plot으로 도출한 각 피처의 중요도를 곱하여 해양오염 지수를 도출하여 새로운 파생 변수 'MPI'를 생성했다.

$$MPI = 0.096468024801441 * x_1 (\text{총인}(\mu\text{g/L})\text{표층}) + 0.0889833550794503 * x_2 (\text{총질소}(\mu\text{g/L})\text{표층}) + \dots + 0.0042507389431043 * x_{35} (\text{날씨})$$

## 2.6. 2차 모델링

### 2.6.1 1년 후 MPI를 나타내는 'yhat' column 생성

해양환경 측정망 데이터로 1년 후를 예측하는 모델을 만들기 위해 1년 후의 MPI를 target으로 설정해주었다. 이는 파이썬의 내장 함수인 shift를 이용해 12개월 후의 행을 대입했다.

### 2.6.2 설명 변수와 목적 변수

1년 후 MPI를 예측하는 회귀 모델은 설명 변수와 목적 변수 간의 함수적인 관계를 학습하여 새로운 입력 값에 대한 출력 값을 예측하는 것이다. 이러한 함수적인 관계를 회귀선 또는 회귀곡선으로 나타낼 수 있다. 이번 AI 모델링에서 회귀 모델의 설명 변수(Feature)와 목적 변수(Target)은 다음과 같이 설정했다.

### 2.6.3 모델 선택 및 평가

회귀 모델의 일반적인 알고리즘에는 선형 회귀(Linear Regression), 랜덤 포레스트(Random Forest) 등이 있다. 그리고 회귀 모델의 평가 지표는 R2 score, MSE(Mean Square Error, 평균 제곱 오차), MAE(mean absolute error, 평균절대오차), SMAPE(Symmetric mean absolute percentage error, 대칭 평균 절대 백분율 오차) 등이 있다.

본 연구의 결과 분석에서 yhat의 증감비율을 활용하기 때문에 이와 같은 단위인 퍼센트(%)를 사용하는 SMAPE에 주목했다. 평균 절대 백분율 오차(MAPE)는 MAE에서 실제 관측값과 회귀식으로 예측한 값 차이를 회귀식으로 예측한 값으로 나눈 절대값의 합이다. SMAPE의 식은 다음과 같다.

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|F_i - A_i|}{(|A_i| + |F_i|)/2}$$

관측 구량 g/L	암모니 아성질소 (μg/L)	암모니 아성질소 (μg/L)	아질산 성질소 (μg/L)	아질산 성질소 (μg/L)	질산성 질소 (μg/L)	질산성 질소 (μg/L)	용존무 기질소 (μg/L)	용존무 기질소 (μg/L)	총질소 (μg/L)	총질소 (μg/L)	용존무 기인 (μg/L)	용존무 기인 (μg/L)	총인 (μg/L)	총인 (μg/L)	규산규 소 (μg/L)	규산규 소 (μg/L)	부유물 질 (mg/L)	부유물 질 (mg/L)	클로로 필 A (μg/L)	WQI 등급	MPI	yhat
2.05	4.20	4.98	2.63	2.63	160.61	164.26	167.44	171.88	264.96	287.64	29.06	32.85	29.98	33.66	332.00	328.36	49.80	24.20	0.85	2.0	16.874618	17.479176
1.54	5.07	5.96	4.30	4.30	299.78	226.25	309.15	236.52	323.95	310.00	40.35	51.15	41.54	57.67	367.30	367.02	22.30	53.30	0.84	2.0	18.436619	17.352797
2.10	3.43	3.49	3.77	3.77	190.08	188.43	197.27	195.68	316.74	281.15	29.20	28.61	38.72	33.20	371.14	361.96	25.40	36.70	0.12	2.0	17.691317	18.496774

Fig. 4. A data frame containing yhat(MPI one year later) columns.

Table 4. Features and Target in a Regression Model

Feature	Target(1년 후 MPI)
36개	회귀
관측년도, 관측월, 날씨, 수심, 투명도(m), 수온(°C)표층, 수온(°C)저층, 염분표층, 염분저층, 수소이온농도표층, 수소이온농도저층, 용존산소량(mg/L)표층, 용존산소량(mg/L)저층, 화학적산소요구량(mg/L)표층, 화학적산소요구량(mg/L)저층, 암모니아성질소(μg/L)표층, 암모니아성질소(μg/L)저층, 아질산성질소(μg/L)표층, 아질산성질소(μg/L)저층, 질산성질소(μg/L)표층, 질산성질소(μg/L)저층, 용존무기질소(μg/L)표층, 용존무기질소(μg/L)저층, 총질소(μg/L)표층, 총질소(μg/L)저층, 용존무기인(μg/L)표층, 용존무기인(μg/L)저층, 총인(μg/L)표층, 총인(μg/L)저층, 규산규소(μg/L)표층, 규산규소(μg/L)저층, 부유물질(mg/L)표층, 부유물질(mg/L)저층, 클로로필A(μg/L)표층, WQI등급, MPI	실수

### 3. 결과 및 고찰

Random Forest 모델을 이용해 train과 test set을 test size 0.2로 분리하여 Confusion Matrix을 도출해본 결과 다음과 같았다.

따라서 Random Forest 모델의 정확도가 약 90%, 평균 교차 검증 정확도가 약 82.61% 이므로 유의미한 모델인 것으로 판별했다.

#### 3.1 연안오염총량관리 오염물질 대상 항목 선정 의사결정 문제 방향성 제안

##### 3.1.1 1차 모델링 결과

모델 선택 과정에서 4가지의 머신러닝 모델을 활용해 성능을 평가해 보았다.

Table 5. Comparison for performance of four ML Models

	Logistic Regression	Random Forest	XGBoost	LGBMClassifier
Hyperparameters	C = 10, penalty = l2, max_iter = 500	max_depth=100, max_features='log2', min_samples_leaf=2, min_samples_split=5, n_estimators=50	n_estimators=100	learning_rate =0.1, n_estimators=1000, colsample_bytree = 0.8
Accuary Score	0.8103	0.8261	0.8363	0.8359

Table 6. Random Forest Model performance results

	Precision	Recall	f1-score	Support
0	0.91	0.98	0.94	1179
1	0.88	0.64	0.74	321
Accuracy			0.90	1500
Macro avg	0.89	0.81	0.84	1500
Weighted avg	0.90	0.90	0.90	1500

#### 3.1.2 각 특별관리해역별 해양오염원의 영향도 도출 : Summary plot



모델이 예측한 특별관리해역일 확률: 0.84 (84%)  
특별관리해역으로 선정하는데 영향이 높은 해양오염원: 아질산성질소표층, 투명도

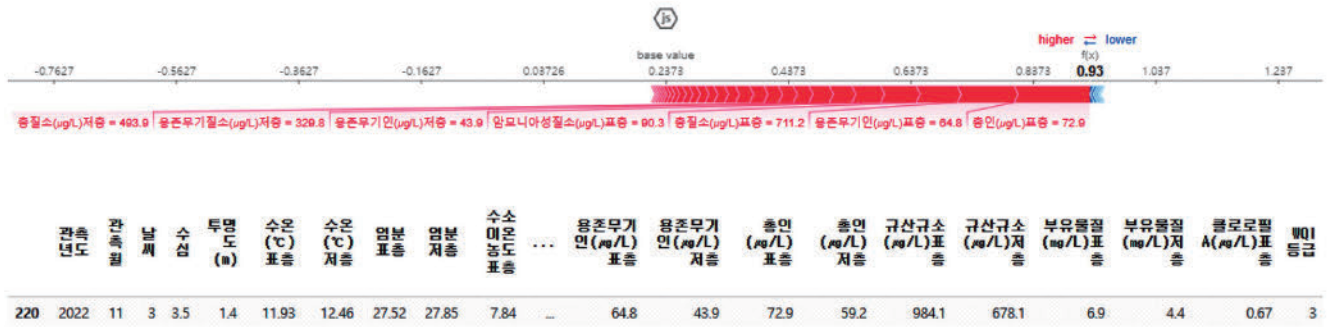
Fig. 5. Special management coastal zone 1) Busan 1.



모델이 예측한 특별관리해역일 확률: 0.94(94%)  
특별관리해역으로 선정하는데 영향이 높은 해양오염원: 총인표층, 암모니아성질소표층

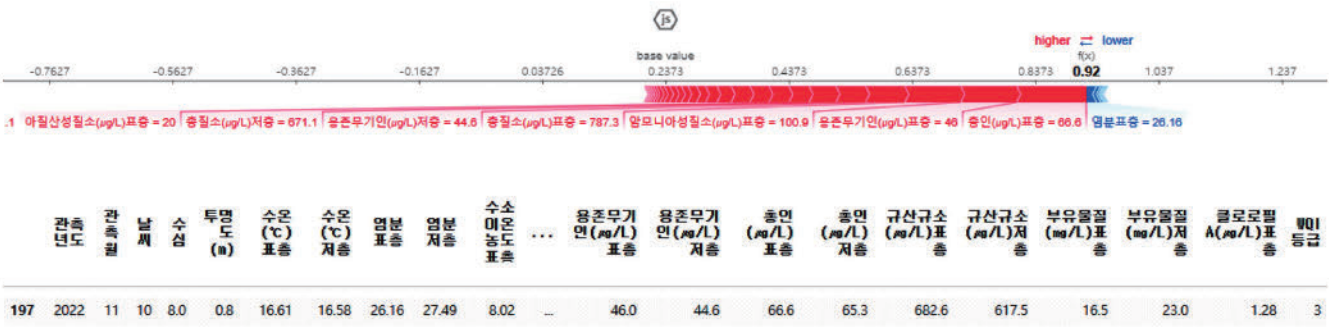
Fig. 6. Special management coastal zone 2) Masan Bay 1.





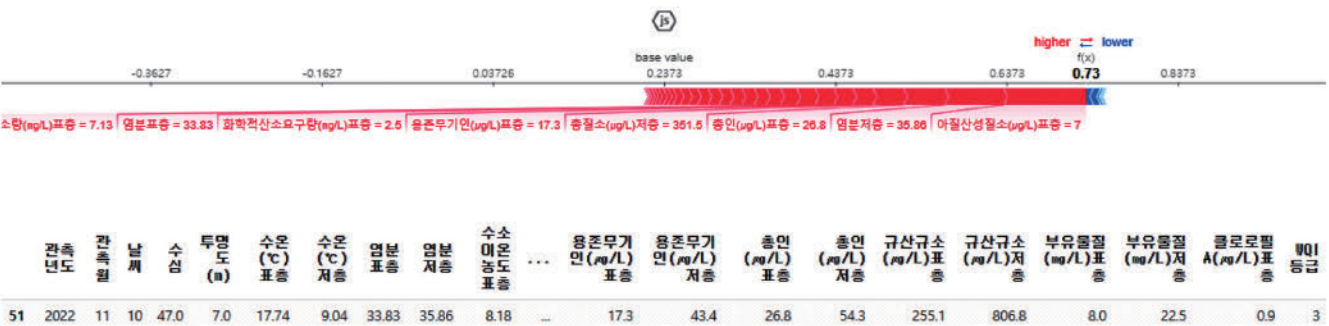
모델이 예측한 특별관리해역일 확률: 0.93(93%)  
 특별관리해역으로 선정하는데 영향이 높은 해양오염원: 총인표층, 용존무기인표층

Fig. 7. Special management coastal zone 3) Shihua Lake 1.



모델이 예측한 특별관리해역일 확률: 0.92(92%)  
 특별관리해역으로 선정하는데 영향이 높은 해양오염원: 총인표층, 용존무기인표층

Fig. 8. Special management coastal zone 4) Incheon 1.



모델이 예측한 특별관리해역일 확률: 0.73(73%)  
 특별관리해역으로 선정하는데 영향이 높은 해양오염원: 아질산성질소표층, 염분저층

Fig. 9. Special management coastal zone 4) Ulsan 1.



3.2 향후 특별관리해역 후보지 선정과 같은 정책 수립 방안 모색

MPI 회귀식을 생성하기 위한 각 피처의 특성 중요도는 다음과 같다.

3.2.1 각 해양오염원 특성 중요도 도출 : Force plot

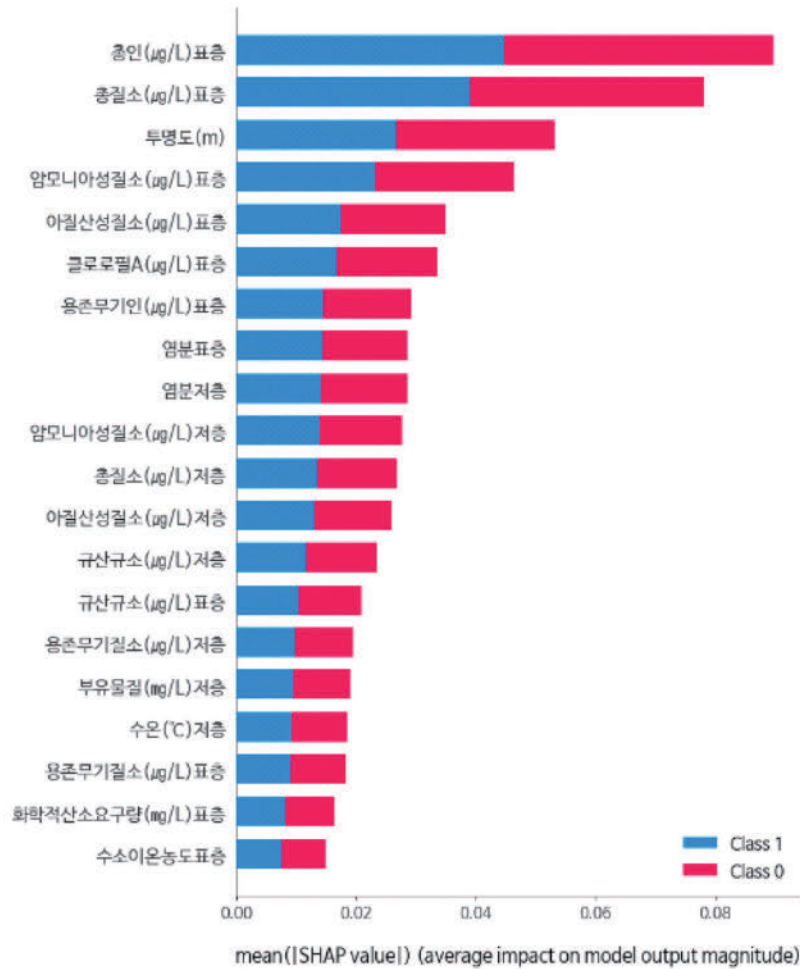


Fig. 10. Force plot feature importance.

Table 7. Attribute importance numbers for each feature

피처1	총인(µg/L)표층	총질소(µg/L)표층	투명도(m)	암모니아성질소(µg/L)표층	클로로필A(µg/L)표층
중요도	0.096468024801441	0.0889833550794503	0.0619450939490201	0.0554878550385068	0.0483941619834996
피처	총질소(µg/L)저층	아질산성 질소(µg/L)표층	암모니아성질소(µg/L)저층	염분저층	용존무기인(µg/L)표층
중요도	0.0452737795926698	0.0409440946049936	0.0400123562801915	0.0381867758707336	0.0352276241632071
피처	아질산성질소(µg/L)저층	염분표층	규산규소(µg/L)표층	용존무기질소(µg/L)표층	규산규소(µg/L)저층
중요도	0.032127368906183	0.0317013057306014	0.0280431429753741	0.0277054564938618	0.0248906031785129
피처	부유물질(mg/L)저층	용존무기질소(µg/L)저층	수온(°C)저층	화학적산소요구량(mg/L)표층	수소이온농도저층
중요도	0.0232982766594573	0.021915502344794	0.0204834411065519	0.0190066716696746	0.0188362743564809
피처	수온(°C)표층	질산성질소(µg/L)표층	총인(µg/L)저층	용존무기인(µg/L)저층	용존산소량(mg/L)저층
중요도	0.0187103310413049	0.018136032512559	0.0179222770205847	0.0164519764699297	0.0156729911685417
피처	화학적산소요구량(mg/L)저층	부유물질(mg/L)표층	수소이온농도표층	질산성질소(µg/L)저층	수심
중요도	0.0155353193060728	0.0155016041565071	0.0147310835420104	0.0137658647767218	0.0133917024780263
피처	용존산소량(mg/L)표층	관측월	WQI등급	관측년도	날씨
중요도	0.0126715332475735	0.0110094784774422	0.0067026443703893	0.0066152577040255	0.0042507389431043

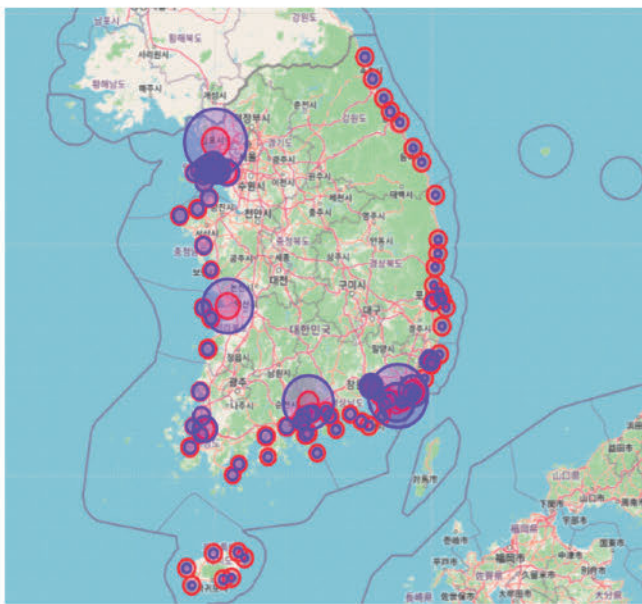


3.2.2 MPI와 해양오염원 상관관계를 도출해 시각화

피어슨 상관계수를 이용하여 MPI와 상관관계가 가장 높은 피쳐를 도출했다. 그 결과 총질소( $\mu\text{g/L}$ )표층이 상관 계수가 0.774916으로 가장 높아 연안의 구역별로 시각화하여 총질소( $\mu\text{g/L}$ )표층과 MPI가 비슷한 분포를 띄는지 시각적으로 파악해 보았다.

	MPI
MPI	1.000000
총질소( $\mu\text{g/L}$ )표층	0.774916
총인( $\mu\text{g/L}$ )표층	0.758998
총질소( $\mu\text{g/L}$ )저층	0.755777
용존무기질소( $\mu\text{g/L}$ )표층	0.731381
용존무기질소( $\mu\text{g/L}$ )저층	0.724862
암모니아성질소( $\mu\text{g/L}$ )표층	0.678539
아질산성질소( $\mu\text{g/L}$ )표층	0.668012
용존무기인( $\mu\text{g/L}$ )표층	0.656848

Fig. 11. Pearson's correlation coefficient figure (Above)



빨간색 원 : MPI,  
 보라색 원 : 총질소( $\mu\text{g/L}$ )표층

Fig. 12. Total nitrogen surface and MPI mapping (Below)

3.2.3 2차 모델링

ML 기법 중 Random Forest 모델을 활용해 train과 test set을 test size 0.2로 분리를 하고 Cross\_val\_score를 사용해 5-fold 교차 검증을 수행결과 다음 Table 8과 같았다.

Table 8. Performance of the quadratic model 2

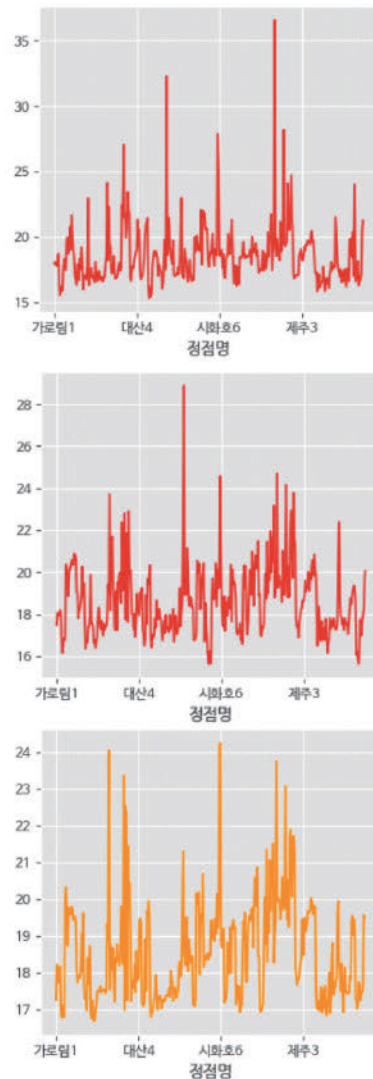
	R2 (결정계수)	평균 교차 검증 R2	MSE	MAE	SMAPE
Score	0.7123	0.7113	2.1913	0.8721	4.1620

모델의 스코어는 R2는 0.71, MAE는 0.87, MSE는 2.2, SMAPE는 4.16%을 확인했다. SMAP 평가지표를 사용한 이유는 0이나 0에 근접한 데이터가 포함되는 경우엔 MAPE는 전체 오류율을 왜곡시키므로 해당 모델에서는 SMAPE가 더 적합한 지표로 사용된다.

따라서 이 평가지표를 통해 모델이 타당성을 판별했고 모델의 설명력을 나타내는 R2가 71% 인 것 뿐만 아니라 SMAPE가 4.16%를 나타내어 유의미한 모델이라는 것을 확인했다.

3.2.4 예측 추세 시각화

실제 구역별 MPI와 예측한 MPI 추세를 시각화 해보았다.



위(빨간선 그래프) : 2021, 2022 2월 실제 해양오염지수  
 아래(주황선 그래프) : 2023 2월 예측 해양오염지수

Fig. 13. Actual vs. predicted MPI trend graph.



3.2.5 1년 후 해양오염지수 증감 상위 10개 해역 선정

향후 특별관리해역 후보지 선정과 같은 정책 수립 방안 모색으로

실제 2022년 2월과 예측 2023년 2월의 해양오염지수 증감 비율 확인하여 상위 10개 해역을 선정하였다.

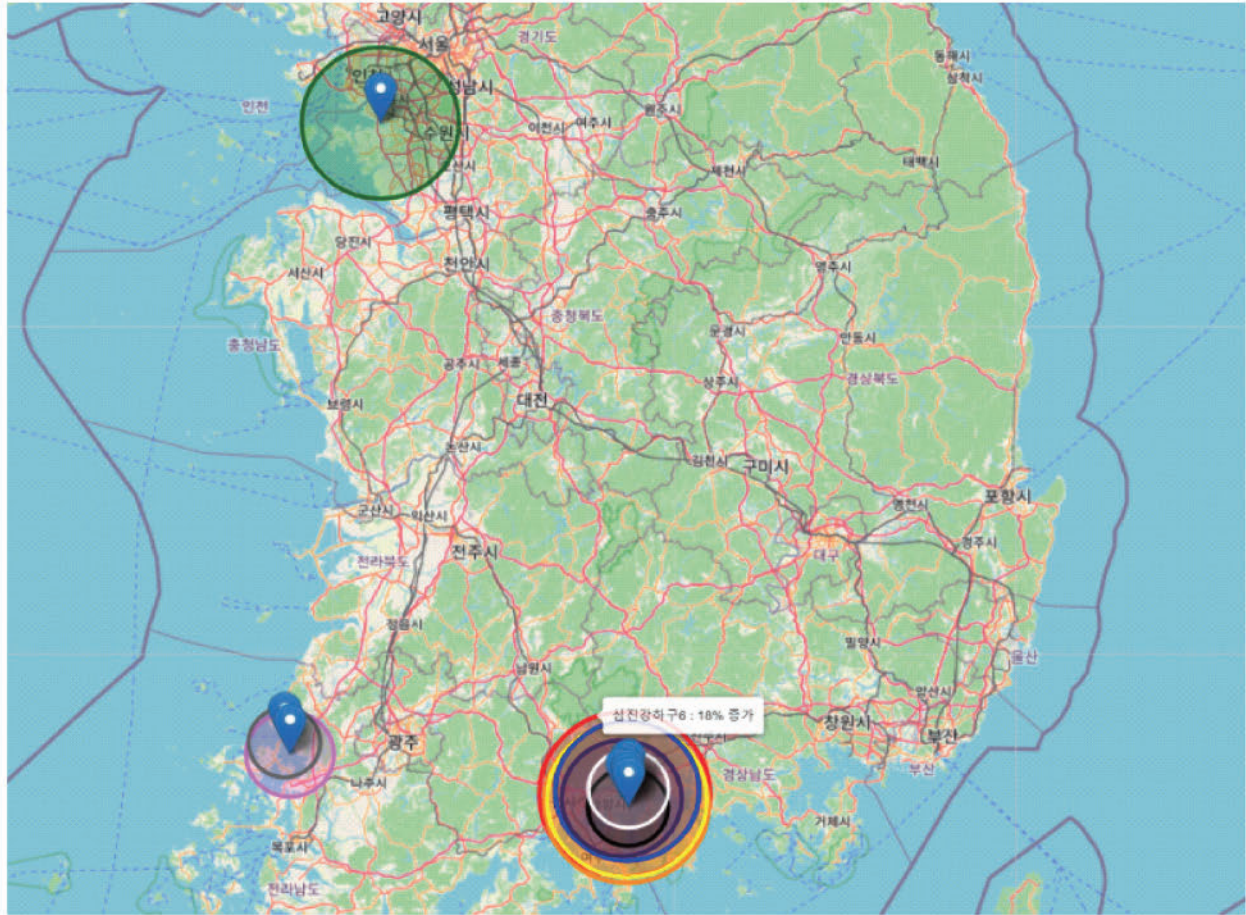


Fig. 14. Visualize the top 10 coastal zones for MPI increase and decrease.

Table 9. Percentage increase or decrease in MPI and latitude and longitude of TOP10 coastal zones

정점명	22년도~23년도 증감 비율(%)	위도(°)	경도(°)
섬진강하구6	18.486268	34.97222	127.76250
섬진강하구9	18.066257	34.94083	127.77111
섬진강하구8	17.374226	34.95278	127.77222
시화호5	16.604971	37.29333	126.75611
섬진강하구5	15.637060	34.98361	127.77306
섬진강하구7	13.007639	34.95778	127.75056
함평4	9.646000	35.11167	126.38389
섬진강하구10	8.678656	34.92667	127.77639
섬진강하구4	8.656328	34.99000	127.77778
함평1	7.670154	35.15278	126.36111

3.3 미래지향적 수질 관리 해결방안 도출

3.3.1 1년 후 MPI 모델링에 대한 XAI

XAI를 활용해 선정된 상위 10개 해역의 해양오염원 도출할 것

이다. 예시로 2022년 2월 1년 후인 2023년 2월 MPI에 가장 영향을 끼치는 해양오염원을 Force\_plot을 활용해 도출하였다.

Table 10. Marine pollution sources by MPI impact in the top 10 coastal zones

정점명	MPI 영향이 높은 해양오염원
섬진강하구6	염분 표층, 총인 저층
섬진강하구9	염분 표층, 총인 저층
섬진강하구8	염분 표층, 총인 저층
시화호5	염분 표층, 총질소 표층
섬진강하구5	염분 표층, 총인 저층
섬진강하구7	염분 표층, 총인 저층
함평4	염분 표층, 총질소 표층
섬진강하구10	염분 표층, 총인 저층
섬진강하구4	염분 표층, 총인 저층
함평1	염분 표층, 총질소 표층

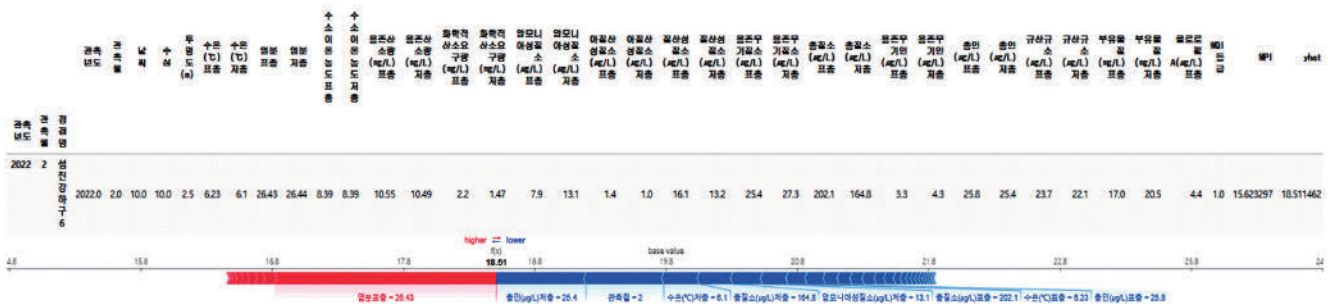


Fig. 15. TOP1) Sumjin river estuary 6 Force\_plot.

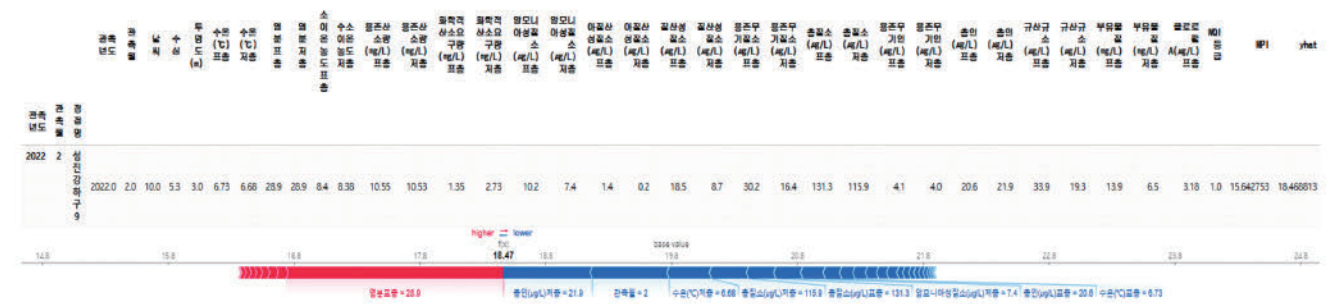


Fig. 16. TOP2) Sumjin river estuary 9 Force\_plot.

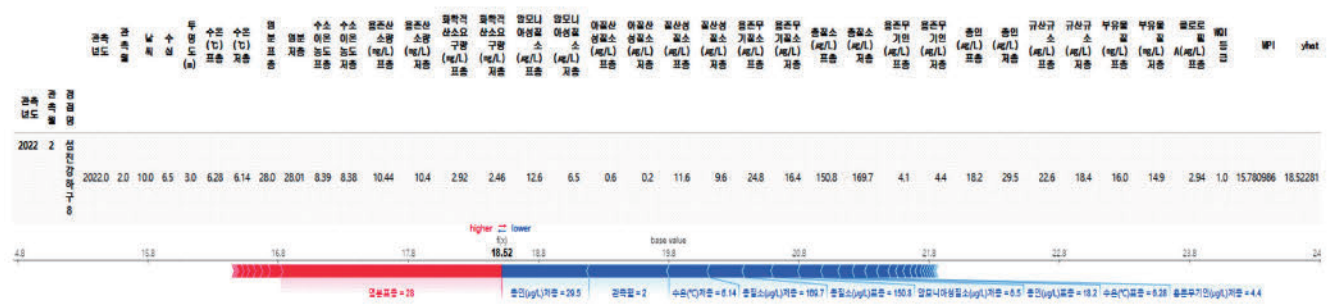


Fig. 17. TOP3) Sumjin river estuary 8 Force\_plot.



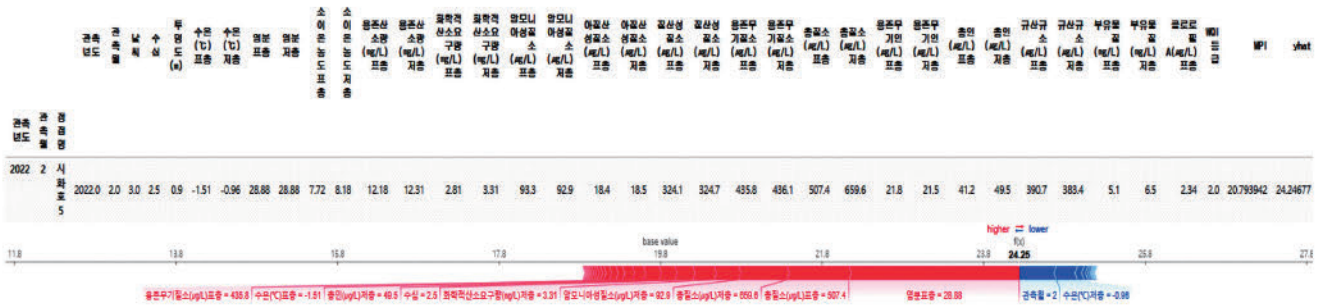


Fig. 18. TOP4) shihua arc 5 Force\_plot.

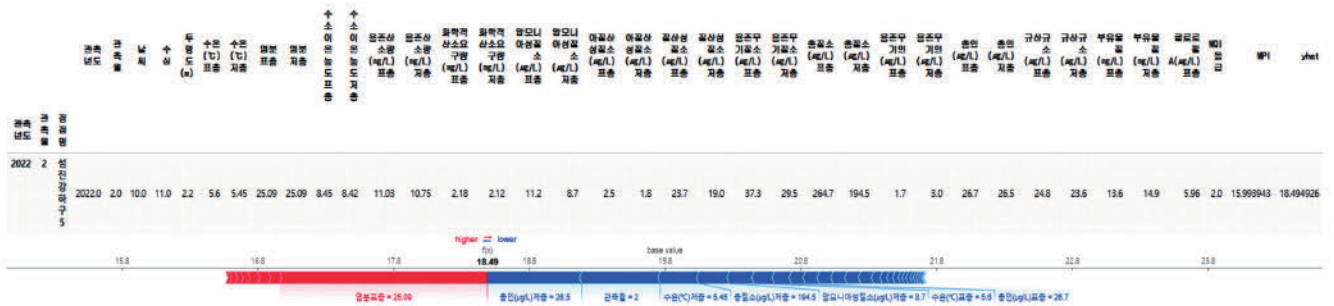


Fig. 19. TOP5) Sumjin river estuary 5 Force\_plot.

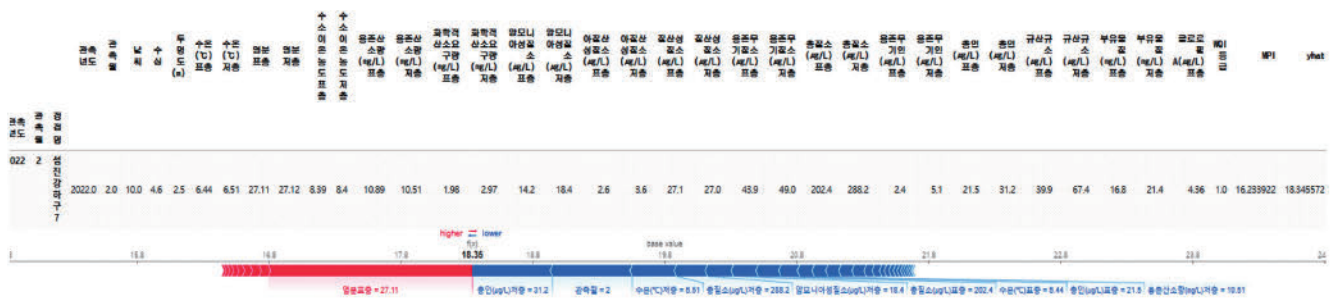


Fig. 20. TOP6) Sumjin river estuary 7 Force\_plot.

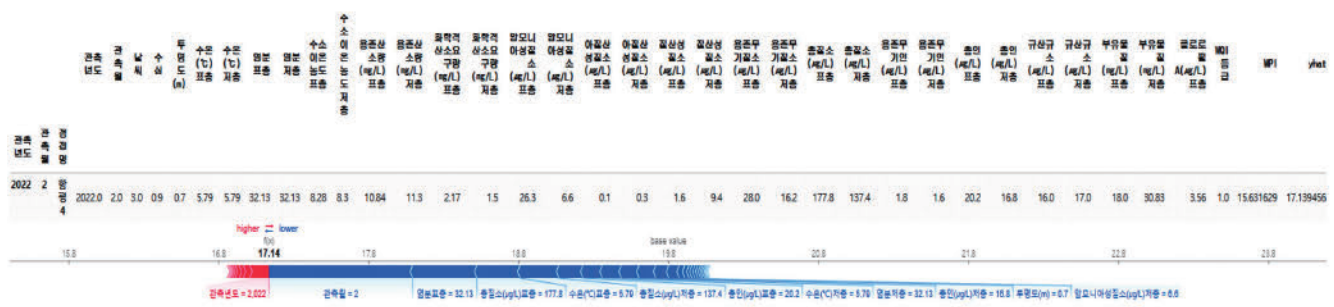


Fig. 21. TOP7) Hampyeong 4 Force\_plot.

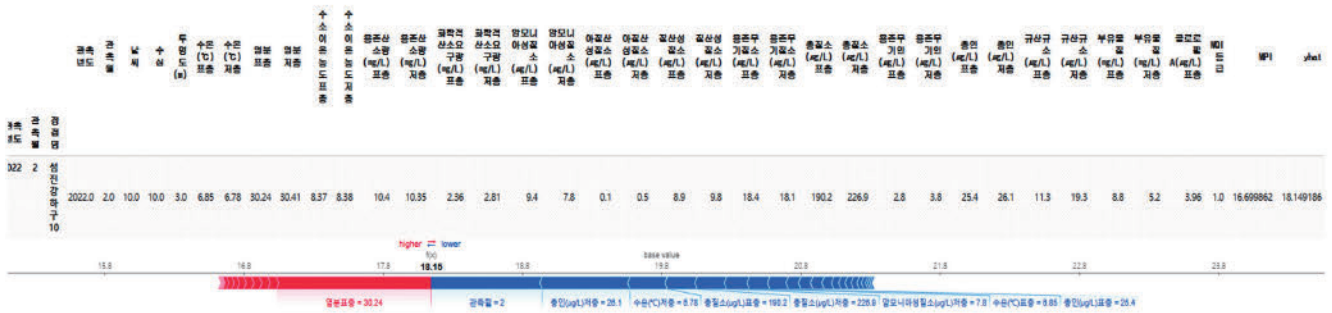


Fig. 22. TOP8) Sumjin river estuary 10 Force\_plot.

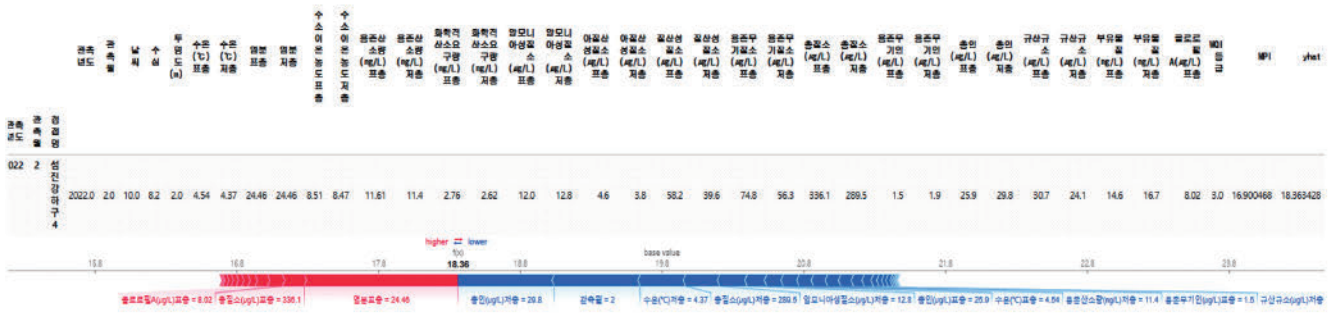


Fig. 23. TOP9) Sumjin river estuary 4 Force\_plot.

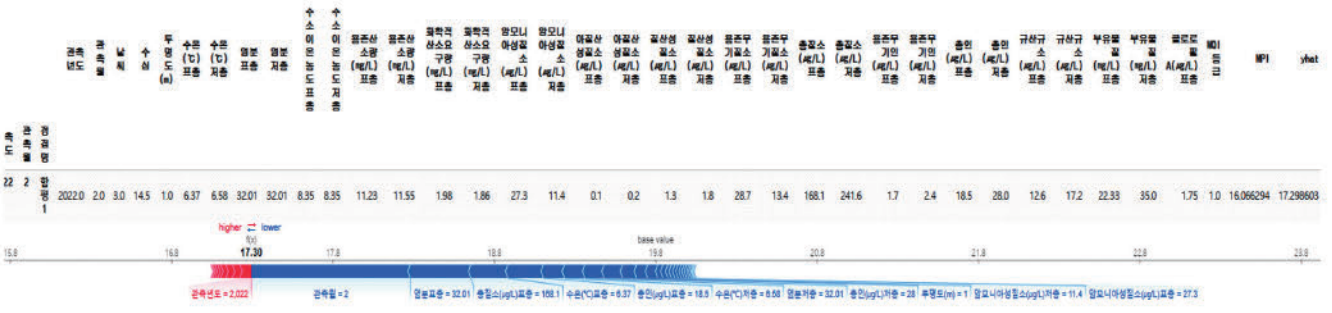


Fig. 24. TOP10) Hampyeong 1 Force\_plot.

3.3.2 미래지향적 수질 관리 해결방안 도출 예시 : 염분표층  
 섬진강하구6 예시로, ‘염분표층’이 1년뒤(2023년 2월) 해양오염  
 지수를 증가시키는 가장 큰 요인으로 도출되었으므로, Fig. 25의  
 summary\_plot 그래프에서 염분표층이 낮아질수록 SHAP value 값  
 이 높아지므로 염분표층이 낮을 수록 1년뒤 해양오염지수가 높아

지는 것을 확인해볼 수 있다. 따라서 섬진강하구6에서 염분표층을  
 높이기 위한 방안을 미리 마련하여 수질 관리를 해야 한다고 도출  
 할 수 있다.



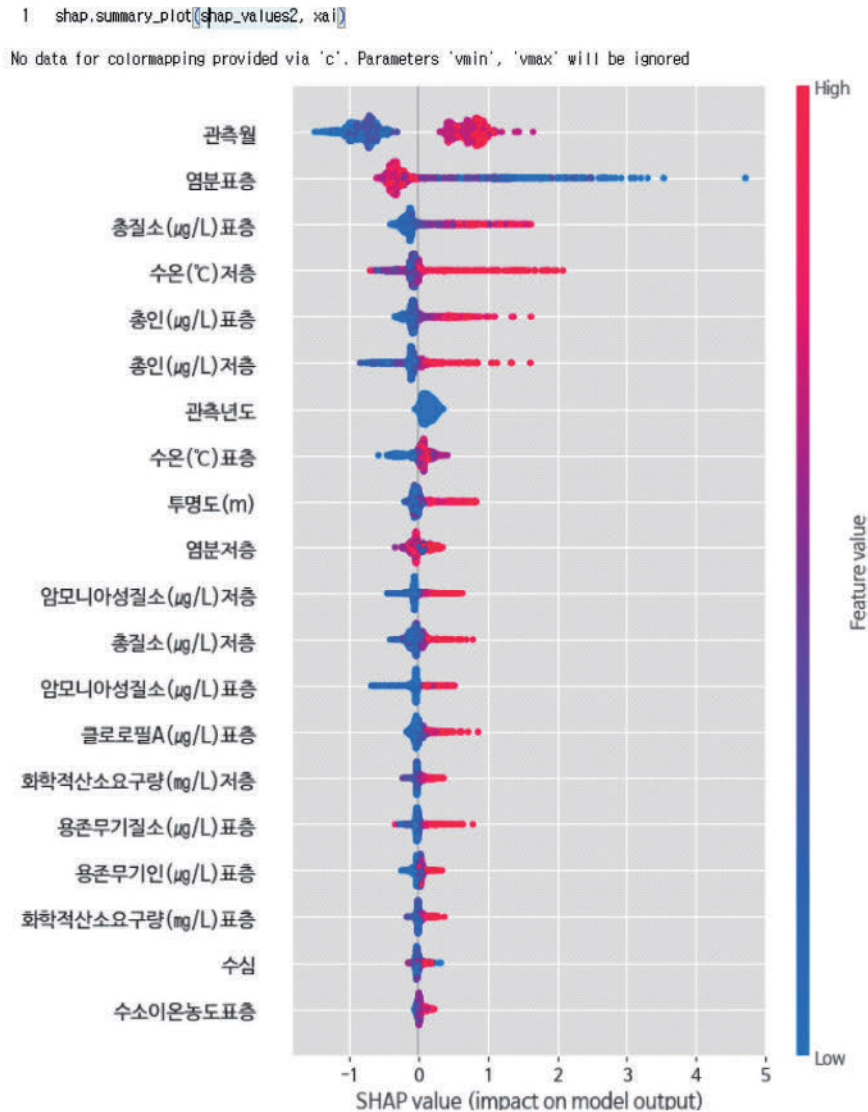


Fig. 25. MPI summary\_plot.

### 3.4 고찰

실험의 XAI를 사용하기 위해 현재 특별관리해역을 타겟으로 하는 ML 기법의 모델링을 사용했다. 연안오염총량관리의 관리대상 항목에서 실제로 ‘총인’이 관리대상물질로 선정 되었다. 그리고 실험의 SHAP 특성 중요도에서 특별관리해역의 연안오염물질로 높은 영향을 끼치는 것으로 선정된 것으로 보아 XAI 결과, 총인이 가장 높게 나온 것을 확인할 수 있다. 이는 규제 항목의 관리에 따른 결과라고 판단할 수 있다. 하지만 연안오염총량관리의 또 다른 규제 오염 물질인 COD는 SHAP의 특성 중요도를 확인해본 결과, COD 저층 29위, COD 표층 35위로 하위권을 차지하였다. 이를 통해 특별관리 해역 선정에서 총인은 높은 영향을 끼치고 있지만 COD는 영향도가 매우 낮다는 것을 확인할 수 있었다.

첫 번째 모델링 XAI의 결과로 회귀식을 생성해 해양오염지수인 MPI를 도출했지만 실제로 이 MPI가 유의미하게 해양오염을 대표하는 지수인지 검증하는 과정이 필요해 보인다. 이러한 과정에서

실험이 진행되어 상위 열 군데 도출한 구역에서 실제 마지막 예시의 섬진강하구6가 염분표층이 실제로 염분표층 때문에 해양오염 관리가 필요한지 현장 조사 과정이 필요해 보인다.

본문에서는 특별관리해역의 각 첫 번째 구역만 대표적으로 나타내었지만 특별관리해역의 각 세부 해역(부산에서는 부산1, 부산2, ... 부산15)마다 영향을 끼치는 해양오염원들이 다르다. 이는 데이터가 더 잘게 쪼개어 수집될수록 더욱 세세하게 관리할 수 있음을 나타낸다.

이 연구에서는 인공지능 모델의 어느 정도 성능(정확도 0.8 이상)만 나타내면 유의미한 모델이라고 판단하여 모델 성능 향상에 대한 고려가 크게 반영되어 있지 않다. 하지만 모델의 성능을 높이기 위한 방법으로 AutoML 혹은 고도화된 딥러닝 기법을 활용해 인공지능 모델링을 고려한다면 0.9 이상의 정밀한 모델이 나올 것이라고 기대된다.



## 4. 결 론

본 연구에서는 크게 세가지 단계로 나누어져 있다. 첫 번째로는 연안오염총량관리 오염물질 대상 항목 선정 의사결정 문제에 새로운 방향성을 제안한다. 우선 2018년부터 2022년까지 5년의 해양환경측정망 데이터를 수집하고 현재 특별관리해역인지에 대해 ML 모델을 이용하여 AI 예측을 하였다. 그리고 예측 과정에서 XAI SHAP 기법을 활용해 각 특별관리해역과 그 해역의 세부 구역에서 어떤 해양오염원이 가장 큰 영향을 끼쳤는지 정밀하게 도출해 보았다. 이 과정에서 XAI 활용하여 force plot, summary plot을 해석하며 특별관리해역의 연안오염총량관리 오염물질에 대하여 새로운 시각으로 접근하며 항목을 선정할 수 있었다.

두 번째로는 향후 특별관리해역 후보지 선정과 같이 인공지능을 활용해 해양오염에 대한 정책 수립 방안을 모색할 수 있었다. 각 해양오염원의 중요도를 도출해 해양오염지수(MPI) 회귀식을 생성하였고 이를 활용하여 ML 모델링을 하여 1년 후 MPI를 예측했다. 예측된 결과를 토대로 1년 후 증감비율을 확인하여 상위 10개 해역을 선정하였다. 이 과정을 통해서 인공지능을 해양오염에 접목하며 추후에 해양오염 관리가 필요한 해역을 선정하여 이를 특별관리해역과 같이 관리가 필요한 해역으로 선정하는 방안을 모색할 수 있었다.

마지막으로 해양오염 관리가 필요한 해역들을 XAI를 활용해 분석하며 해역별로 해양오염지수에 영향을 끼친 해양오염원을 도출해 보았다. 이를 통해 예측한 해양오염원의 관리 방안을 모색할 수 있어 미래지향적이고 장기적인 효과를 추구하는 수질 관리 해결방안을 도출하였다.

## 후 기

본 논문은 해양환경측정망 생산 정보의 해석평가와 활용방법을 찾기 위하여, 해양수산부에서 주최하고 해양환경공단 및 한국해양환경 에너지학회에서 주관하는 “2023년 해양환경측정망 자료 활용 논문 공모전”에 참가하기 위해 수행된 연구임.

## References

- [1] Sun, H.J., Yoo, S.J., Noh, J.S. and Yoo, P.J., 2009, Bioactive Carbon Process in Water Treatment, *Journal of the Korean Society of Environmental Engineers*, 31(4), 308-323.
- [2] Lee, Y.S., Yu, J.S., Jeong, S.Y. and Choi, Y.R., 2003, Bacillus sp. Ammonia and Nitrite Nitrogen Removal in Water Quality by A8-8, *Journal of Life Sciences*, 13(1), 47-53.
- [3] Jeon, E.J., Lee, Y.M., Lee, D.I. and Kim, G.Y., 2018, Current status of use and development of environmental management sea areas, trends in water quality change, and improvement plan for impact assessment, *Journal of the Korean Society of Marine Environment & Safety*, 726-728.
- [4] Community Advisory Council for busan castal area, What is a Specially Management Coastal Zone?, <http://www.busancoast.or.kr/ko/7>, 2019.
- [5] Ministry of Oceans and Fisheries, Introduction of Marine Environmental Standards: Water Quality Index, <https://www.meis.go.kr/mei/wqi/introduce.do>, 2018.
- [6] Ministry of Oceans and Fisheries, Basic Policy for Total Pollutant Load Management System in Special Management Coastal Zone, <https://www.law.go.kr/admRulLsInfoP.do?admRulSeq=2100000071152#AJAX>, 2016.
- [7] Ministry of Land, Infrastructure and Transport, Implementation of Total Pollutant Load Management System, [https://www.molit.go.kr/USR/policyData/m\\_34681/dtl?id=509](https://www.molit.go.kr/USR/policyData/m_34681/dtl?id=509), 2012.
- [8] Ministry of Oceans and Fisheries, 84% of coastal waters nationwide have a water quality rating of ‘good’, [https://www.mof.go.kr/statPortal/bbs/report/view.do?ntt\\_id=691&pageIndex=&searchType=&searchQuery=](https://www.mof.go.kr/statPortal/bbs/report/view.do?ntt_id=691&pageIndex=&searchType=&searchQuery=), 2017.
- [9] Kangjun Lee Consultant, Traditional data analysis methodology: KDD, CRISP-DM, <https://www.2e.co.kr/news/articleView.html?idxno=301010>, 2020.
- [10] Ministry of Environment (Green Transition Policy Division), Framework Act on Environmental Policy - National Law Information Center, <https://www.law.go.kr/%EB%B2%95%EB%A0%B9%ED%99%98%EA%B2%BD%EC%A0%95%EC%B1%85%EA%B8%B0%EB%B3%B8%EB%B2%95>, 2023.
- [11] Emedia editorial office, need to correct the trajectory of the total phosphorus elimination policy, <http://www.ecomedia.co.kr/news/newsview.php?ncode=179507196329468>, 2012.
- [12] Park, I.G., There is no improvement in Saemangeum water quality without resolving the stratification caused by salinity, <https://www.hani.co.kr/arti/area/honam/907128.html>, 2019.
- [13] e-Country Indicator, Status of Water Quality (COD) in Specially Management Coastal Zone, [https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx\\_cd=1273](https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1273), 2022.
- [14] KMA, Maritime Numerical Forecast Weather Map, [https://www.weather.go.kr/mini/marine/roms\\_saln.jsp?type=M&data=glosea\\_post\\_grph\\_nwpacific\\_salt\\_&ftm=F000&depth=000](https://www.weather.go.kr/mini/marine/roms_saln.jsp?type=M&data=glosea_post_grph_nwpacific_salt_&ftm=F000&depth=000), 2023.
- [15] Ministry of Education, Change in Salinity, <https://m.blog.naver.com/PostView.naver?isHttpsRedirect=true&blogId=moe-blog&logNo=220489584592>, 2015.
- [16] Park, D.H., Explanation of Shapley value, SHAP, Tree SHAP, <https://daehani.tistory.com/entry/Shapley-value-SHAP-Tree-SHAP>, 2020.
- [17] Lee, G.H., Logistic regression parameter estimation, <https://ratsgo.github.io/machine%20learning/2017/07/02/logistic/>, 2017.

Received 11 September 2023

Revised 13 November 2023

Accepted 14 November 2023