

# XGBoost를 활용한 연근해어업 어획량 예측 연구

김지웅 · 정지훈\*

국립부경대학교(교수) · \*국립부경대학교(학생)

## A Study on Predicting Coastal Fishery Catches Using XGBoost

Ji-Ung KIM · Ji-Hoon JEONG\*

Pukyong National University(professor · \*student)

### Abstract

This study compared the predictive performance of a traditional multivariate time series model, the Vector Autoregression model, and a machine learning-based XGBoost model for the mid to short-term forecasting of coastal fishery catches. The results showed that while the VAR model, which accounts for seasonal volatility and lag effects in historical time series data, exhibited relatively high predictive volatility, the XGBoost model demonstrated more stable predictions by learning nonlinear patterns and complex inter-variable relationships. Notably, in the XGBoost model, lagged variables of past catches and derived variables related to aquaculture production were identified as important predictors, confirming that coastal fishery catches are significantly influenced by short-term volatility and aquaculture production factors. Although both models showed similar overall trend directions, differences were observed in the magnitude of fluctuations. The VAR model generally indicated a declining trend, replicating past patterns in its forecasts, while the XGBoost model showed a more gradual and stable decline.

**Key words :** Coastal fishery, Capture production, XGBoost, VAR

### I. 서론

우리나라 연근해 어업자원은 기후변화뿐만 아니라 배타적 경제수역(EEZ) 설정, 과잉 어획, 연안개발, 해양환경 변화, 어선 감척 등 복합적인 영향으로 변동해왔다(Kim and Kang, 2025). 연근해어업 어획량은 1960년대 이후 어선 척수 증가, 동력화 및 규모화로 1996년까지 120만 톤~170만 톤 규모를 유지했으나 이후 어족자원 감소로 하락세로 돌아섰다. 연근해어업 어획량은 역사적으로 2016년 44년만에 처음으로 100만 톤 미만으로

하락했으며, 이후 감소 추세가 이어져 2024년 84만 톤을 기록 중이다(KOSIS, 2025).

연근해어업 주요 어획 어종은 시대별로 해양 환경에 변화에 따라 달라지고 있다. 상위 어획 어종은 1970년대 오징어, 갈치, 멸치, 1980년대 쥐치, 멸치, 갈치, 1990년대 쥐치, 정어리, 멸치, 2000년대 오징어, 멸치, 고등어, 2010년대 멸치, 오징어, 고등어 등이다. 향후 수온 상승에 따라 연근해 어업생산량은 감소할 것이라는 연구가 다수 보고된 바 있다(Kim et al., 2024).

연근해어업 어획량 변화는 양식업과도 밀접한

\* Corresponding author : 051-629-5963, abcd7138@naver.com

\* 본 연구는 2025년도 교육부 및 부산시의 재원으로 부산RISE혁신원의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다.(2025-RISE-02-001-006)

관계를 가지고 있다. 해면양식업은 연근해어업과 동일한 어장을 공유하고 있어 수산자원뿐만 아니라 환경적인 측면을 공유하고 있다. 양식업의 확대는 수산물 생산 구조의 재편과 함께 연근해어업의 상대적 생산 비중 변화로 이어졌다. 우리나라 양식업은 1960~1970년대 흑색혁명(미역, 다시마 등 해조류)과 굴 양식의 확산을 시작으로 1980년대 굴, 홍합, 김, 1990년대 넙치, 조피볼락 등 활어 양식 대중화, 2000년대 전복 양식 등이 발달하면서 기르는 어업으로의 전환을 일궈냈다(NIFC, 2021). 양식업이 발달함에 따라 많은 어업인들이 양식업으로 업종을 전환하는 구조적 변화도 이어져 왔다. 현재 양식업은 국내 수산물 생산(원양 제외) 중 약 70%, 연근해어업은 약 30% 비중을 차지하고 있다(KOSIS, 2025)

수요 측면에서 연근해 수산물은 세계적인 교역 확대로 빠르게 변화하고 있다. 수산물 공급은 1997년 수입 자유화 이후 수입 수산물 비중이 증가하고 있다. 수입산 수산물은 고등어, 낙지, 조기 등 대중어종뿐만 아니라 연어, 새우, 바닷가재 등 고가격 수산물로 확대되고 있다(Kim and Jang, 2016). 연근해어업은 글로벌화로 인해 세계 수산물 시장은 긴밀하게 연결되고 있으며, 이로 인해 직간접적으로 영향을 받고 있다.

향후 연근해어업은 지속적으로 쇠퇴할 것으로 전망되고 있으나, 정확한 어획량 예측이 어려운 실정이다. 연근해어업은 1994년부터 어선감축사업을 통해 자원대비 과도한 어선세력을 구조조정하려는 노력을 이어오고 있으나 지속적이고 체계적인 진행이 이루어지지 못한다는 지적이 있다(Lee and Heu, 2018). 향후 과학적인 어획량 예측은 적정 어선세력 규모, 톤수, 단위노력당 어획량(CPUE) 등 연근해어업에 투입되는 노력량 조절에 중요한 통찰을 제공할 수 있다는 점에서 중요하다.

기존 선행연구에서는 ARIMA 등 시계열 모델을 통해 연근해어업 어획량을 추정해왔다(Park and Yoon, 1996; Cho et al., 2006; Kwak, 2020). 그러나 기존 예측 방법은 비선형적 관계와 복잡

한 환경 변수 간의 상호작용을 충분히 반영 못하는 한계가 존재한다. 본 연구에서는 기존 시계열 모델을 보완할 수 있는 XGBoost(eXtreme Gradient Boosting)를 사용하여 어획량을 예측하였다. XGBoost 모델은 다양한 유형의 복합 데이터를 통합 처리할 수 있으며, 여러 개의 의사 결정나무를 조합해 사용하는 예측 모델로 비선형 및 상호작용의 자동 학습을 통해 높은 예측 정확도를 기대할 수 있다(Jeong, 2023).

본 연구에서는 연근해어업, 양식업, 수출입 부문 변수와 XGBoost, VAR 모형 활용하여 향후 연근해어업 어획량을 추정하고자 했다.

## II. 연구 방법

### 1. 선행연구 검토

Kwak(2020)은 ARIMA 모형과 전이함수 모형을 활용해 주요 어종(고등어, 조기, 갈치, 멸치, 오징어)의 어획량을 예측하고, 해양환경 요인이 어획량 변동에 미치는 영향을 분석하였다. 어획량, 수온(서해, 동해, 남해), 소비자 물가지수(어류 및 수산) 자료를 활용한 분석 결과, 동해 수온은 조기 어획량에, 남해·동해 수온은 고등어 어획량에 유의한 영향을 미치는 것으로 나타났다. 이는 전통적 시계열 모형에서도 외생 환경요인이 어획량 예측에 중요함을 시사한다.

Hyun and Cho(2021)은 다변량 시계열 모형인 VAR모형을 적용하여 광어 생산량을 예측하고, 경쟁 어종과의 상호관계를 분석하였다. 광어와 연어, 조피볼락, 방어, 참돔 간 관계를 분석한 결과, 연어만이 광어 생산량에 유의한 영향을 미치는 것으로 나타났으며, 생산량 예측 결과는 2025년까지 연간 약 41천 톤 수준으로 유지될 것으로 분석되었다. 해당 연구는 단변량 접근을 넘어 어종 간 상호작용을 고려한 시계열 예측의 필요성을 제시한다.

Nam and Noh(2012)은 다중회귀분석, SARIMA,

VECM 모형을 활용하여 굴의 월별 생산량 단기 예측을 수행하고 모형 간 예측 성능을 비교하였다. 굴 생산량 및 위판가격, 수온(통영), 강수량 자료를 활용한 결과, 다중회귀모형이 SARIMA와 VECM모형에 비해 실제값과 예측값 간 오차가 가장 작은 것으로 나타났다. 이는 단기 예측에서는 비교적 단순 선형 모형의 우수한 예측력을 보여준다.

Nam et al.(2014)은 김 실질 위판가격을 대상으로 다중회귀모형, ARIMA, VECM 모형을 적용하여 단기 예측 성능을 비교하였다. 김 산지가격, 수출량, 1인당 양곡 소비량, 콩치 생산량, 수온 등 여러 시차 변수를 활용한 분석 결과, 다중회귀모형이 ARIMA(1,0,0) 및 VECM 모형보다 높은 적합도를 보였다. 연구는 가격 예측에서 다변량 설명변수를 포함한 회귀 접근의 우수한 적합도를 제시한다.

Lee and Kim(2022)은 머신러닝 기반의 LSTM 과 GRU 모델을 활용하여 양식산 조피볼락의 산지가격을 예측하고, 모델별 및 투입된 변수 개수에 따른 예측력을 평가·비교하였다. 양식산 조피볼락의 지역별 산지가격(통영, 여수, 완도), 출하량, 양성 물량, 지역별 도매가격(인천, 하남, 부산), 수출입 물량 등 19가지 변수를 사용하였다. 분석 결과, GRU보다 LSTM 모델의 예측력이 우수하게 나타났으며, 단변량 모델보다 다변량 모델의 예측 성능이 높은 것으로 분석되었다.

Lim and Kim(2025)은 CNN-LSTM with Attention 모델을 활용하여 멸치 어획량을 장기적으로 예측하고, 불법 조업 방지 및 어업 관리 활용 가능성을 제시하였다. 수온, 풍속, 기온, 파고 등 해양환경 자료와 과거 어획량 데이터를 결합한 분석 결과, 사용된 모델은 전반적으로 계절성과 장기 추세를 동시에 반영하는 안정적인 예측 성능이 확인되었다. 이는 딥러닝 기반 모형이 장기 어획량 예측에 적용될 경우 안정적인 예측 성능을 보일 수 있음을 시사한다.

Park et al.(2024)은 장단기 기억 신경망(LSTM)

모형을 적용하여 인천 해역 꽃게 어획량을 예측하고, 해양환경 변수의 포함 여부에 따른 예측 정확도를 비교하였다. 월별 인천광역시 꽃게 어획량, 인천광역시 해역의 평균 수온, 강수량, 풍속, 파고, 파주기 자료를 사용하였다. 분석 결과, 과거 어획량 자료만을 학습한 경우보다 해양환경 자료를 함께 학습한 경우 예측 정확도가 향상되는 것으로 나타났다. 이는 기계학습 모형에서 외생 환경 변수의 활용이 예측력 개선에 기여할 수 있음을 보여준다.

Cho et al.(2006)은 ARIMA와 SARIMA 모형을 활용하여 주요 4개 어종(멸치, 고등어, 갈치, 참조기)의 어획량을 예측하였다. 해당 연구는 영향요인과의 관계 분석보다는 시계열 자료 자체의 특성을 반영한 예측 모형 도출에 초점을 두었다. 분석 결과, SARIMA 모형이 ARIMA 모형에 비해 상대적으로 안정적인 예측 성능을 보였다.

Kim and Kim(2021)은 머신러닝 알고리즘인 XGBoost와 Ridge Regression, Random Forest 모형을 활용하여 작물 생산량 예측의 적합도를 비교·분석하였다. 스마트팜 빅데이터를 활용하였으며, 작기별 정보 서비스 데이터와 Provide 서비스 데이터로 구분되어있다. 분석 결과, XGBoost 모형이 최적의 예측 성능을 보였으며, 비선형적 관계를 학습하는 기계학습 모형의 예측 우수성이 확인되었다. 이는 복잡한 생산 결정 구조를 갖는 자료에서 기계학습 모형의 활용 가능성을 제시한다.

제시된 선행연구들은 분석 모델 측면에서 시계열 전통 모형(ARIMA, SARIMA, VAR, VECM)부터 머신러닝 및 딥러닝 모델(LSTM, GRU, XGBoost, CNN-LSTM)까지 다양한 방법론을 적용하여 수산물 및 농작물의 생산량 예측을 수행하고 있으며, 최근 연구일수록 인공지능 기법을 도입하는 경향을 보이고 있다. 연구 주제 및 영향요인 측면에서는 단순한 시계열 예측을 넘어 해양환경요인(수온, 강수량 등), 경쟁 어종 관계, 시장 변수(가격, 소비량, 무역량) 등 다양한 외생변수를 통해 예측 성능을 높이는 것으로 나타났다.

## 2. 분석 자료

본 연구에서 분석 자료는 2010년 1월부터 2024년 12월까지의 월별 시계열 데이터를 활용하였다. 주요 변수는 종속변수인 연근해어업 총 어획량(톤)이며, 독립변수로 양식 총 생산량, 수출금액, 수입금액, 어선 척수, 어선 톤수, 마력수, 어업인구 수를 포함하였다. 자료는 통계청(KOSIS), 한국수산무역협회, 한국해양수산개발원(KMI)에서 제공하는 통계 데이터를 활용하였다.

예측 기간은 2025년 1월부터 2027년 12월까지의 36개월이며, 이 기간동안의 어획량 변화를 단·중기적으로 예측하는 것을 목표로 하였다.

<Table 1> Analysis Data

Contents	
Raw data	2010.01~2024.12 (180 months)
Forecast period	2025.01~2027.12 (36 Months)
Method	VAR XGBoost

<Table 2> Definition of Variables

Variables		Unit	Definition of Variable	Source	References
Dependent	Coastal Catch	ton	Total amount of costal fishery production	KOSIS	Nam and Noh (2012)
	Aquaculture Production	ton	Total amount of aquaculture production	KOSIS	Hyun and Cho(2021), Nam and No(2012)
	Export Value	USD (1,000)	Export value of fishery products, excluding processed products	KITA	Nam and Noh (2012)
	Import Value	USD (1,000)	Import value of fishery products, excluding processed products	KITA	Lee and Kim (2022)
Independent	Vessel Count	vessels	Total number of registered fishing vessels	KOSIS	Lee and Heu (2018)
	Vessel Tonnage	ton	Total gross tonnage of fishing vessels	KOSIS	Lee and Heu (2018)
	Horsepower	hp	Total engine horsepower of fishing vessels	KOSIS	Lee et al. (2006)
	Fishery Population	persons	Total population of fishery households	KMI	Moo et al. (2019)

## 3. 분석 방법

본 연구에서는 연근해 어업 어획량 예측을 위해 전통적 다변량 시계열 모형인 VAR(Vector Autoregression) 모형과 기계학습 기반의 XGBoost 모형을 적용하였다. 두 모형은 동일한 자료와 예측 기간에 적용하여 예측 결과를 비교·분석함으로써, 선형 시계열 구조와 비선형 패턴 학습 간 예측 특성의 차이를 검토하고자 한다.

### 가. VAR 모형

VAR모형은 다변량 시계열 자료에서 변수 간의 선형적 상호작용과 시차 효과를 동시에 반영할 수 있는 모형으로(Moon, 1997), 수식 (1)과 같이 정의된다.

$$Y_t = C + \sum_{i=1}^p A_i Y_{t-i} + \epsilon_t \dots\dots\dots (1)$$

여기서  $Y_t$ 는 연근해 어획량, 양식 생산량, 어선 규모 등으로 구성된  $k \times 1$ 차원의 내생변수 벡터를 의미하며,  $A_i$ 는 시차  $i$ 에서의 계수 행렬,  $p$ 는 시차 차수,  $\epsilon_t$ 는 백색잡음 오차항이다.

모형 추정에 앞서 모든 변수에 대해 ADF(Augmented Dickey-Fuller) 검정을 실시하여 시계열의 정상성을 확인하였다. 일부 변수에서 단위근이 존재하는 것으로 나타나 1차 차분 또는 로그 변환을 적용하였으며, 변환된 변수를 VAR 모형 추정에 활용하였다. 최적 시차는 AIC(Akaike Information Criterion)를 기준으로 설정하였고, 2010년 1월부터 2024년 12월까지의 자료를 이용해 모형을 추정하였다.

추정된 VAR 모형을 바탕으로 연근해어업 어획량에 대한 예측을 수행하였으며, 2025년 1월부터 2027년 12월까지의 기간을 대상으로 연속적 예측(rolling forecast) 방식으로 예측값을 산출하였다. 이를 통해 어획량의 선형적 자기상관 구조와 변수 간 시차 효과가 반영된 예측 결과를 도출하였다. 다만 VAR 모형은 선형 구조를 전제로 하므로, 계절성 변화나 외생 변수 간의 복합적 상호작용과 같은 비선형적 특성을 충분히 반영하는 데에는 한계가 있을 수 있다.

나. XGBoost 모형

본 연구에서는 비선형적 특성 반영의 필요성을 고려하여, VAR 모형과 함께 XGBoost 모형을 병행 적용하였다. XGBoost는 결정트리 기반의 부스팅 알고리즘으로, 변수 간의 비선형적 관계와 복합적인 상호작용을 효과적으로 학습할 수 있는 기계학습 모형이다. XGBoost의 목적함수는 수식 (2)와 같이 정의된다.

$$obj(\theta) = \sum_i^n l(y_i; \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \dots\dots\dots (2)$$

XGBoost의 목적함수는 실제값  $y_i$ 와 예측값  $\hat{y}_i$  간의 오차를 측정하는 손실함수와, 개별 결정트리  $f_k$ 의 복잡도를 제어하는 규제항으로 구성된다. 이를 통해 XGBoost 모형은 예측 오차 최소화와 과적합 방지를 동시에 고려한 예측을 수행한다.

본 연구에서는 시계열적 특성을 반영하기 위해

각 변수에 대해 1, 3, 6, 12개월의 시차(lag)변수를 생성하였으며, 단기 변동성을 포착하기 위해 이동평균(rolling mean) 및 표준편차(rolling std) 변수를 추가하였다. 또한 계절성을 고려하기 위해 월(month)과 분기(quarter) 변수를 추가하고, 월 변수는 순환형 특성을 반영하여 sin-cos 변환을 통해 인코딩하였다.

모형 학습은 시계열 순서를 유지한 상태에서 2010~2021년을 학습 구간, 2022~2023년을 검증 구간, 2024년을 테스트 구간으로 분할하여 수행하였다. 학습에서는 회귀 손실함수를 사용하였으며, 주요 하이퍼파라미터는 TimeSeriesSplit 기반 교차검증과 Randomized Search를 통해 최적화하였다. XGBoost 모형은 2025년 1월부터 2027년 12월까지의 어획량을 연속적 예측(rolling forecast) 방식으로 추정하였다.

### Ⅲ. 연구 결과

#### 1. VAR 모형

본 연구에서는 연근해 어획량의 변동을 동태적으로 분석하고 향후 추세를 예측하기 위해, 2010년 1월부터 2024년 12월까지의 월별 시계열 자료를 바탕으로 다변량 시계열 모형인 VAR(Vector Autoregression) 모형을 구축하였다. 분석 모형의 종속변수는 ‘연근해 총 어획량’으로 설정하였으며, 이를 설명하는 독립변수로는 대체재 성격의 ‘양식 생산량’, 시장 수요 요인인 ‘수출금액’과 ‘수입금액’, 그리고 생산 노력 투입 요소인 ‘어선 척수’, ‘어선 톤수’, ‘마력수’, ‘어업인구수’를 포함하여 총 8개 변수 간의 상호 의존성을 고려하였다.

모형 추정에 앞서 수행한 단위근 검정(Unit Root Test)결과, 모든 변수에서 단위근이 검출되어 1차 차분을 수행하였다. 차분 후 재검정 결과 모든 변수가 정상성(Stationarity)를 만족하는 것으로 나타나 이를 최종 분석 자료로 활용하였다. 또한 VAR모형의 예측력을 좌우하는 적정 시차

(Lag) 선정은 정보 기준 AIC(Akaike Information Criterion)을 적용하였으며, 검정 결과 최적 시차는 8개월(Lag=8)로 도출되었다. 이는 연근해 어업 생산 활동이 전년도 동기뿐만 아니라, 약 8개월 전의 생산량에도 유의한 영향을 받고 있음을 의미한다.

VAR(8)모형의 적합도를 확인한 결과, 모형의 설명력을 나타내는 결정계수(R<sup>2</sup>)는 0.88로 높게 나타났다. 또한 모형의 적합도 지표인 AIC값은 141.857로 나타났다.

마지막으로 추정된 모형의 안정성을 검증하기 위해 잔차(Residual)의 상관행렬을 분석한 결과(<Table 3>), 대부분의 변수 간 상관계수는 ±0.3 이하로 나타나 모형 내 오차항 간의 상호의존성이 크지 않은 것으로 확인되었다. 이는 VAR 모형의 기본 가정이 충족되었음을 의미하며, 추정 결과의 신뢰성을 뒷받침한다(Sims, 1980).

구축된 VAR모형을 이용하여 2025년 1월부터 2027년 12월까지의 향후 36개월간의 연근해 어획량을 예측한 결과([Fig. 1]), 전반적으로 과거의 계절적 순환 패턴을 반복하며 특정 범위 내에서

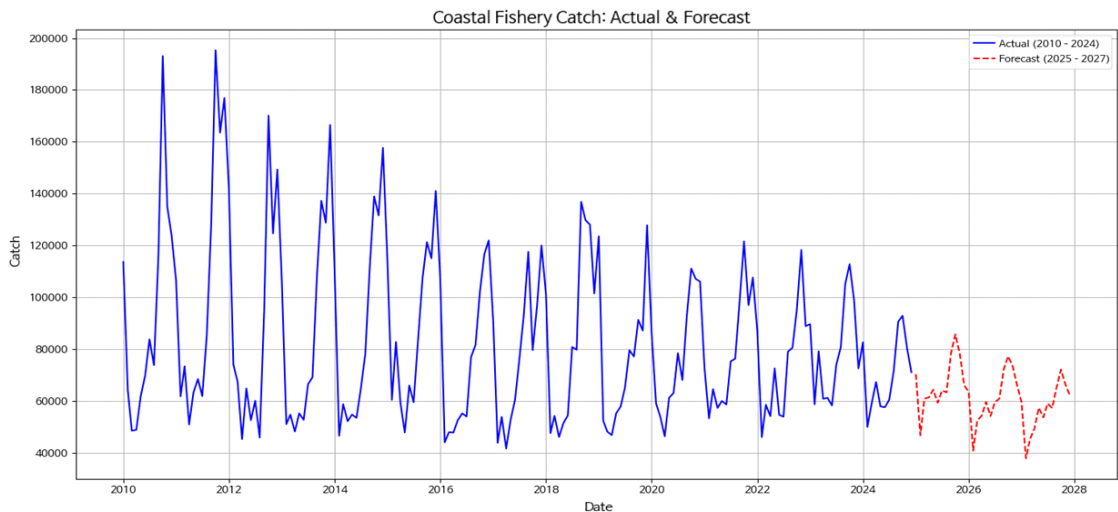
등락을 거듭하는 것으로 나타났다. 예측된 어획량의 추이는 점차 감소하는 추세로 나타났다.

<Table 3> Residual Correlations with Coastal Catch

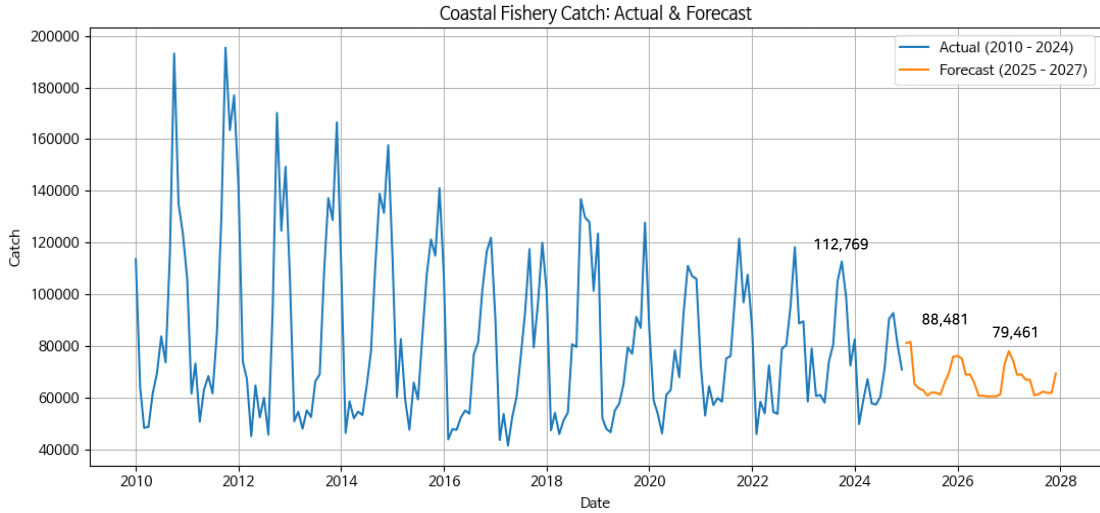
Variable	Correlation
Aquaculture Production	-0.06
Export Value	0.26
Import Value	0.30
Vessel Count	0.04
Vessel Tonnage	-0.08
Horsepower	-0.10
Fishery Population	0.23

## 2. XGBoost 결과

본 연구에서는 전통적 시계열 모형의 선형적 한계를 보완하고, 데이터의 복잡한 비선형적 패턴을 포착하기 위해 기계학습 기반의 XGBoost Regression 모형을 구축하였다. 분석 기간은 동일하게 2010년 1월부터 2024년 12월까지의 월별 자료를 기반으로 연근해 어획량을 예측하였다([Fig. 2]).



[Fig. 1] VAR Model Prediction of Coastal Fishery Catches.

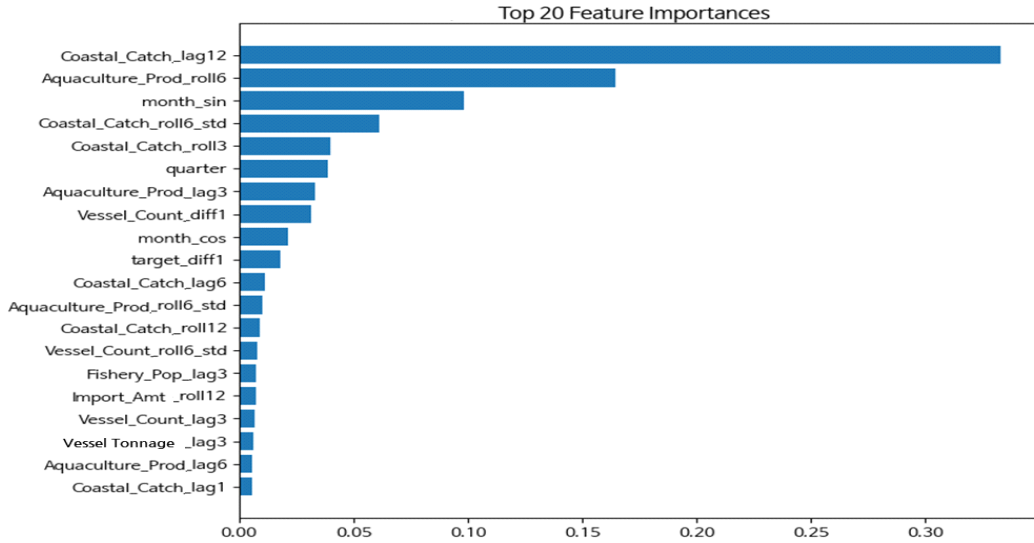


[Fig. 2] XGBoost Model Prediction of Coastal Fishery Catches.

중속변수인 연근해 총 어획량을 설명하기 위해 양식 총 생산량, 수출금액, 수입금액, 어선 척수, 어선 톤수, 마력수, 어업인구 수 등의 다차원적인 독립변수를 투입하였다. 특히, 시계열 데이터의 동태적 특성을 반영하기 위해 주요 변수에 대해 1개월에서 12개월까지의 시차(Lag) 변수와 이동평균(Rolling Mean) 변수를 파생 변수로 생성하였으며, 월별 계절성을 학습시키기 위해 삼각함수 기반의 순환형 인코딩(month\_sin, month\_cos) 변수를 추가하였다.

학습된 XGBoost 모형의 예측 성능을 평가한 결과, 모형의 설명력을 나타내는 결정계수(R<sup>2</sup>)는 0.94로 높게 나타났으며, 예측오차의 크기를 나타내는 RMSE는 3154.36으로 산출되었다. 통상적으로 시계열 예측에서 0.9 이상의 R<sup>2</sup>는 매우 높은 수준의 적합도를 의미한다(Chen and Guestrin, 2016). 이는 XGBoost 알고리즘이 경사 하강법(Gradient Descent)을 통해 이전 트리의 오차를 순차적으로 보정하는 부스팅(Boosting)기법을 사용하여(Chen and Guestrin, 2016), 단순한 선형 관계뿐만 아니라 변수 간의 복잡한 상호작용 효과까지 학습하기 때문이다.

모형 예측 과정에서 각 변수가 기여한 정도를 나타내는 변수 중요도 분석 결과는 [Fig. 3]과 같다. 분석 결과, ‘연근해 어획량\_lag 12’가 가장 높은 중요도를 기록하였다. 이는 전년도 동월의 어획량이 현재 어획량을 결정하는 주요 요인임을 의미한다. 다음으로 ‘양식 생산량\_roll6’변수가 높은 중요도를 보였다. 이는 최근 6개월간의 양식 생산 추세가 연근해 어획량 변동에 밀접하게 연관되어 있음을 의미한다. 양식 생산량의 증감은 수산물 시장 내에서 연근해 어획물의 대체 또는 보완 관계를 형성하여 가격 기제 등을 통해 생산 유인에 영향을 미친 결과로 해석된다. 또한 계절성을 나타내는 ‘month\_sin’ 변수가 상위권에 위치하여, 월별 수온, 해류, 금어기 등의 계절적 요인이 어획량 변동의 영향을 미치는 요인임을 보여준다. 이외에도 단기 변동성을 나타내는 ‘연근해 어획량\_roll6\_std’, ‘양식 생산량\_lag3’, ‘quarter’ 등의 변수가 중요도가 높게 나타났다. 이는 단기적 생산 변동성과 분기별 생산 패턴이 예측력 향상에 기여했음을 의미한다. 반면, 어선 톤수나 어업인구 수와 같은 구조적 변수들의 중요도는 상대적으로 낮게 나타났다.



[Fig. 3] Variable importance of the XGBoost model.

이는 중요하게 작용하였는데, 이는 장기적 산업 기반의 변화가 단기 예측에는 직접적으로 큰 영향을 미치지 않는 것으로 해석된다.

### 3. VAR, XGBoost 모형 비교

본 연구에서는 전통적 다변량 시계열 분석 기법인 VAR모형과 기계학습 기반의 XGBoost모형을 활용하여 2025년 1월부터 2027년 12월까지 향후 36개월간의 연근해 어획량을 예측하였다. 분석 결과, 두 모형 모두 단기적으로 급격한 추세 하락이나 상승 없이 특정한 수준의 어획량을 유지하는 경향을 보였으나, 구체적인 예측 패턴과 수치에서는 모형 간 차이가 확인되었다([Fig. 4]).

VAR 모형은 과거 시계열 데이터가 가진 계절적 변동성을 반영하여 비교적 큰 변동성을 보이는 것으로 나타났다. 성어기인 9월~11월 구간에서 약 72,000~85,000톤 수준의 생산량을 기록할 것으로 예측되었다. 이는 VAR모형이 변수 간의 선형적 상호작용과 과거 시차(Lag)의 영향력을 기반으로 미래를 예측하기 때문으로 해석된다

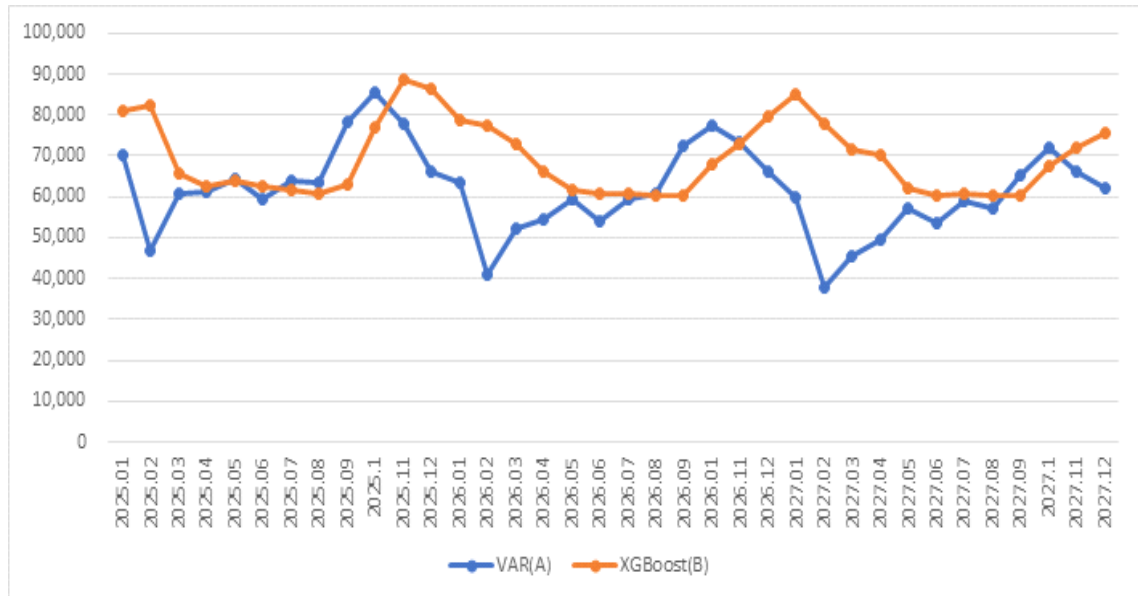
(Hyun JE et al., 2021).

반면, XGBoost 모형은 VAR 모형에 비해 상대적으로 적은 변동 흐름을 보였으며, 동일한 성어기(9월~11월)에 약 82,000~88,000톤 수준을 기록하여 VAR 모형보다 다소 높은 어획량을 예측하였다. 일반적인 시계열 분석에서 머신러닝 기법 중의 하나인 XGBoost 기법은 변수 간의 복잡한 비선형적 상호작용을 학습하여 일반화된 패턴을 도출하는데 강점이 있기 때문이다(Chen and Guestrin, 2016). 결과적으로 XGBoost 모형이 단순한 과거 추세의 반복보다는 데이터 전체의 평균적인 지시선(Bas-line)을 학습했기 때문이다.

결과적으로 두 모형의 예측값 차이는 선형성을 가정하는 통계적 모형(VAR)과 비선형성을 학습하여 복잡한 상호작용을 반영하는 알고리즘 모형(XGBoost)의 구조적 차이에서 나타난 것으로 보인다. VAR모형이 과거의 변동성을 크게 반영하여 예측을 했다면, XGBoost 모형은 학습시 최근의 시계열 데이터 패턴과 변수 간의 비선형적 관계를 반영하여 연근해 어업의 어획량의 잠재력을 긍정적으로 평가한 것으로 해석할 수 있다.

<Table 4> Comparison of VAR and XGBoost Models for Predicting Coastal Fishery Catches (unit : ton)

Time	2025			Time	2026			Time	2027		
	VAR (A)	XGBoost (B)	B-A		VAR (A)	XGBoost (B)	B-A		VAR (A)	XGBoost (B)	B-A
2025.01	70,329	80,863	10,535	2026.01	63,559	78,931	15,373	2027.01	59,679	85,202	25,523
2025.02	46,734	82,442	35,709	2026.02	40,820	77,242	36,422	2027.02	37,841	77,934	40,093
2025.03	60,913	65,480	4,567	2026.03	52,363	72,815	20,451	2027.03	45,243	71,499	26,256
2025.04	61,362	62,422	1,060	2026.04	54,358	66,216	11,858	2027.04	49,395	70,313	20,917
2025.05	64,375	63,918	-457	2026.05	59,597	61,866	2,269	2027.05	57,370	61,996	4,626
2025.06	59,275	62,689	3,414	2026.06	54,145	60,710	6,565	2027.06	53,647	60,472	6,826
2025.07	64,081	61,465	-2,616	2026.07	59,531	60,912	1,381	2027.07	59,025	60,610	1,586
2025.08	63,262	60,821	-2,441	2026.08	60,845	60,431	-414	2027.08	57,239	60,288	3,049
2025.09	78,485	62,839	-15,646	2026.09	72,362	60,358	-12,005	2027.09	65,315	60,358	-4,957
2025.10	85,719	76,723	-8,996	2026.10	77,233	68,011	-9,223	2027.10	72,136	67,704	-4,432
2025.11	78,053	88,481	10,428	2026.11	73,481	73,104	-378	2027.11	66,329	72,210	5,881
2025.12	66,014	86,214	20,201	2026.12	66,226	79,462	13,235	2027.12	62,204	75,628	13,424
Total	798,602	854,357	55,755	Total	734,520	820,058	85,538	Total	685,423	824,214	138,791



[Fig. 4] Comparison of VAR and XGBoost Models for Predicting Coastal Fishery Catches(unit : ton).

## IV. 결론

본 연구는 연근해어업 어획량의 중·단기 예측을 위해 전통적 다변량 시계열 모형인 VAR모형과 기계학습 기반의 XGBoost 모형을 병행 적용하여 예측 결과를 비교하였다.

분석 결과, VAR 모형은 과거 시계열이 지닌 계절적 변동성과 시차 효과를 반영하여 예측 변동성이 상대적으로 크게 나타난 반면, XGBoost 모형은 비선형적 패턴과 변수 간 복합적 관계를 학습함으로써 상대적으로 안정적인 예측이 관찰되었다. XGBoost 모형에서는 과거 어획량의 시차 변수와 양식 생산 관련 파생 변수가 예측 과정에서 중요한 변수로 활용되어, 연근해어업 어획량이 단기적 변동성과 양식업 생산 요인의 영향을 크게 받는 것을 확인할 수 있었다.

특히, 본 연구 XGBoost 모형을 통한 어획량 예측에서는 전년 동월 어획량과 함께 최근 6개월 양식 생산량이 중요 변수로 나타났다. 이는 연근해어업이 반복되는 어기의 계절성에 영향을 받음과 동시에 양식업의 생산량이 직간접적으로 어획량과 관련 있음을 의미한다. 양식업은 연근해어업과 해면 어장을 공유하고 있어 생태적, 시장경제적 측면에서 복합적인 영향을 주는 것으로 볼 수 있다.

두 모형의 예측 결과는 전반적인 추세 방향에서는 유사하였으나, 변동 폭에서는 차이가 나타났다. VAR 모형은 전반적으로 하락하는 추세가 나타났으며, 과거 패턴을 반복하는 예측 흐름이 보였다. XGBoost 모형은 전반적으로 완만하고 안정적인 하락 추세가 나타났다. 연도별 연근해어업 총어획량은 VAR 모형과 XGBoost 모형 모두에서 감소세를 보이는 것으로 분석되었다. VAR 모형은 2025년 798,602톤, 2026년 734,520톤, 2027년 685,423톤으로 예측되어 향후 급속한 감소를 보인 반면, XGBoost 모형은 2025년 854,357톤, 2026년 820,058톤, 2027년 724,214톤으로 상대적

으로 완만한 감소를 보이는 것으로 분석되었다.

분석 결과를 종합할 때, 연근해어업 어획량 예측은 특정 유형의 단일 예측 모형에 의존하기 보다는, 선형적 시계열 구조를 반영하는 통계 모형과 비선형적 패턴 학습에 강점을 지닌 기계학습 모형을 병행하여 활용하는 접근이 효과적임을 시사한다. VAR모형은 어획량 변동의 시계열적 구조를 이해하고 과거 패턴을 반영한 예측에 유용하며, XGBoost 모형은 복잡한 변수 간 관계를 학습하여 예측의 안정성을 제고하는 데 강점을 가진다. 따라서 향후 어획량 예측 연구에 있어서는 예측 목적과 활용 시점에 따라 두 모형을 상호 보완적으로 활용하는 것이 효과적일 것이다.

한편, 본 연구는 다음과 같은 한계를 지닌다. 첫째, 본 연구는 연근해어업 총 어획량을 대상으로 분석을 수행하였기 때문에, 어종별 자원 특성이나 어업 유형별 차이를 충분히 반영하지 못하였다. 어종별로 어획 시기와 자원 변동이 상이하다는 점을 고려할 때, 총량 자료를 활용한 분석은 개별 어종 수준의 예측에 비해 예측 정확도에 한계를 가질 수 있다. 둘째, 본 연구에서 활용한 독립변수는 주로 생산, 어업 여건 관련 지표로만 구성되어 있어, 수온 및 기후 변동성과 같은 해양환경 요인을 포함하지 않았다. 향후 연구에서는 어종별 분석 및 환경 변수 추가를 통해 어획량 예측 연구를 확장할 수 있을 것으로 생각된다. 셋째, 본 연구에서는 VAR모형과 XGBoost 모형의 예측 결과를 비교하기 위해 결정계수와 AIC 등 성능 지표를 함께 제시하였으나, 두 모형은 구조적 가정과 학습방식이 상이하여 설명력 지표를 직접적으로 비교하는 데에는 한계가 존재한다. 이에 본 연구에서는 성능 지표에 따른 해석을 보조적으로 활용하고, 두 모형의 예측 결과가 보이는 전반적인 경향과 변동성 차이를 중심으로 비교 및 해석하였다.

## References

- Chen T(2016). XGBoost: A Scalable Tree Boosting System. Cornell University.
- Cho YJ, Cho YH and Kim JY(2006). Time Series Analysis of Catch in Coastal Sea Fisheries, *Journal of Rural Development*, 29(1), 123~134.
- Hyun JE and Cho JH(2021). Production Prediction of Halibut Using VAR Model : Focusing on the Production and Importance of Competing Fish Species, *Korean Management Consulting Review*, 21(3), 213~221.
- Kim C, Lee JS, Yang JY and Han IS(2024). Dynamic downscaling for regional ocean climate modeling around the Korean peninsula and its application in fisheries. *Korean Journal of Fisheries and Aquatic Sciences*, 57(2), 177~185.  
<https://doi.org/10.5657/KFAS.2024.0177>
- Kim SW and Kim YH(2021). A study on the application of machine learning techniques for crop yield prediction. *Journal of the Korea Academia-Industrial Cooperation Society*, 22(7), 403~408.
- Kim DS and RA HR(2016). The Economic Effects of Oil Tariff Reduction of Korea-GCC FTA based on VAR Model, *International Area Studies Review*, 20(1), 23~51.
- Korean Statistical Information Service(2025). Fisheries Production Ssurvey.  
<https://kosis.kr> on December 18.
- Kwak NY(2020). Impact of Water Temperature on Fish Catch and Fisheries CPI: An Intervention Analysis, *The Statistical Review* 21, 87~100.  
<https://kiss.kstudy.com/Detail/Ar?key=3750125>
- Lim SH, Kim JU and Jang YS(2017). A Study on the Positioning of Sliced Raw Fish Market by Selection Attributes. *The Journal of Fisheries Business Administration* 48(2), 53~66.  
<https://doi.org/10.12939/fba.2017.48.2.053>
- Lee GH and Kim DH(2022). A study on predicting farm-gate prices of aquaculture rockfish using deep learning techniques: Focusing on LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) models. *Journal of Food Distribution Research*, 39(4), 21~41.
- Lee YG, Yu JW, Kim KS and Kang DS(2006). A Study on the Effective Horsepower Estimation for Domestic Coastal Fishing Vessels, *Journal of the Society of Naval Architects of Korea*, 43(3), 313~321.  
<https://doi.org/10.3744/SNAK.2006.43.3.313>
- Lim MS and Kim DY(2025). Anchovy Catch Prediction and Illegal Fishing Prevention Using CNN-LSTM Attention, *Maritime Science & Technology*, 59, 19~42.
- Moon GS(1997). A understanding of Vector Autoregressive Model, *Journal of The Korean Official Statistics*, 2(1), 23~56.
- Mo SW, Lee KB and Park JH(2019). The Factor Decomposition of Fishery Household Populations Using Shift-Share Analysis, 31(1), 101~116.
- Nam JO, Baek EY and Noh SG(2014). Forecast for Laver Producer Price Using Time Series Models, *Ocean and Polar Research*, 29(2), 271~303.  
<http://dx.doi.org/10.35372/kmiopr.2014.29.2.010>
- Nam JO and Noh SG(2012). A Study on Forecast of Oyster Production using Time Series Models, *Ocean and Polar Research*, 34(2), 185~195.  
<http://dx.doi.org/10.4217/OPR.2012.34.2.185>
- National Institute of Fisheries Science(2021).  
<https://www.nifs> on December 18.
- Park MG, Koo CM, and Byun SS(2024). Prediction of Swimming Crab Catch Amount in Incheon Waters Using Long Short-term Memory (LSTM) and Consideration of Its Use in Naval Operations, *Journal of the KNST*, 7(3), 309~315.  
<http://dx.doi.org/10.31818/JKNST.2024.9.7.3.309>
- Sims CA(1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1~48.

- 
- Received : 23 December, 2025
  - Revised : 21 January, 2026
  - Accepted : 28 January, 2026