



# Contents

- 01 | 연구배경 및 필요성
- 02 | 데이터
- 03 | 분석 및 결과
- 04 | 적용과 기대효과

## 01 연구배경 및 필요성

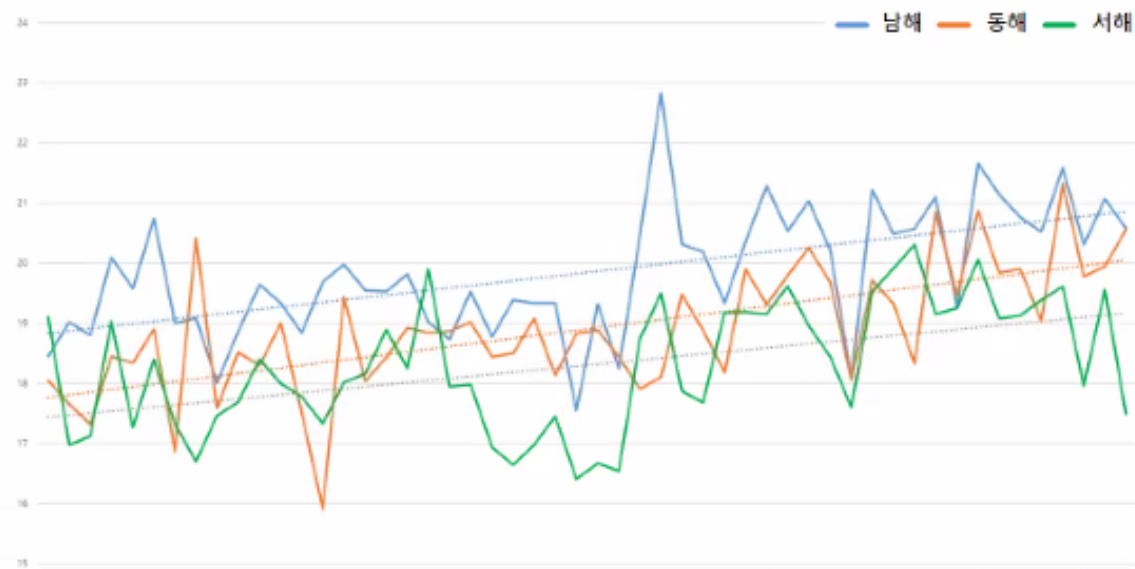
- 한반도 연근해는 다른 해역에 비해 수온 변화에 민감
- 각 해역별로 서해 1.5°C, 남해 2°C, 동해 2.4°C 상승

〈 전세계 우리나라의 상승 표층 수온 비교 〉

	상승 표층 수온(°C)		
	전 세계(A)	우리나라(B)	비교 (C=B/A)
최근 50년 (1968~2017년)	<b>0.52</b>	<b>1.12</b>	<b>2.2(배)</b>
최근 30년 (1988~2017년)	0.31	0.29	1.0(배)

출처\_국립수산물과학원 「기후변화연구과」

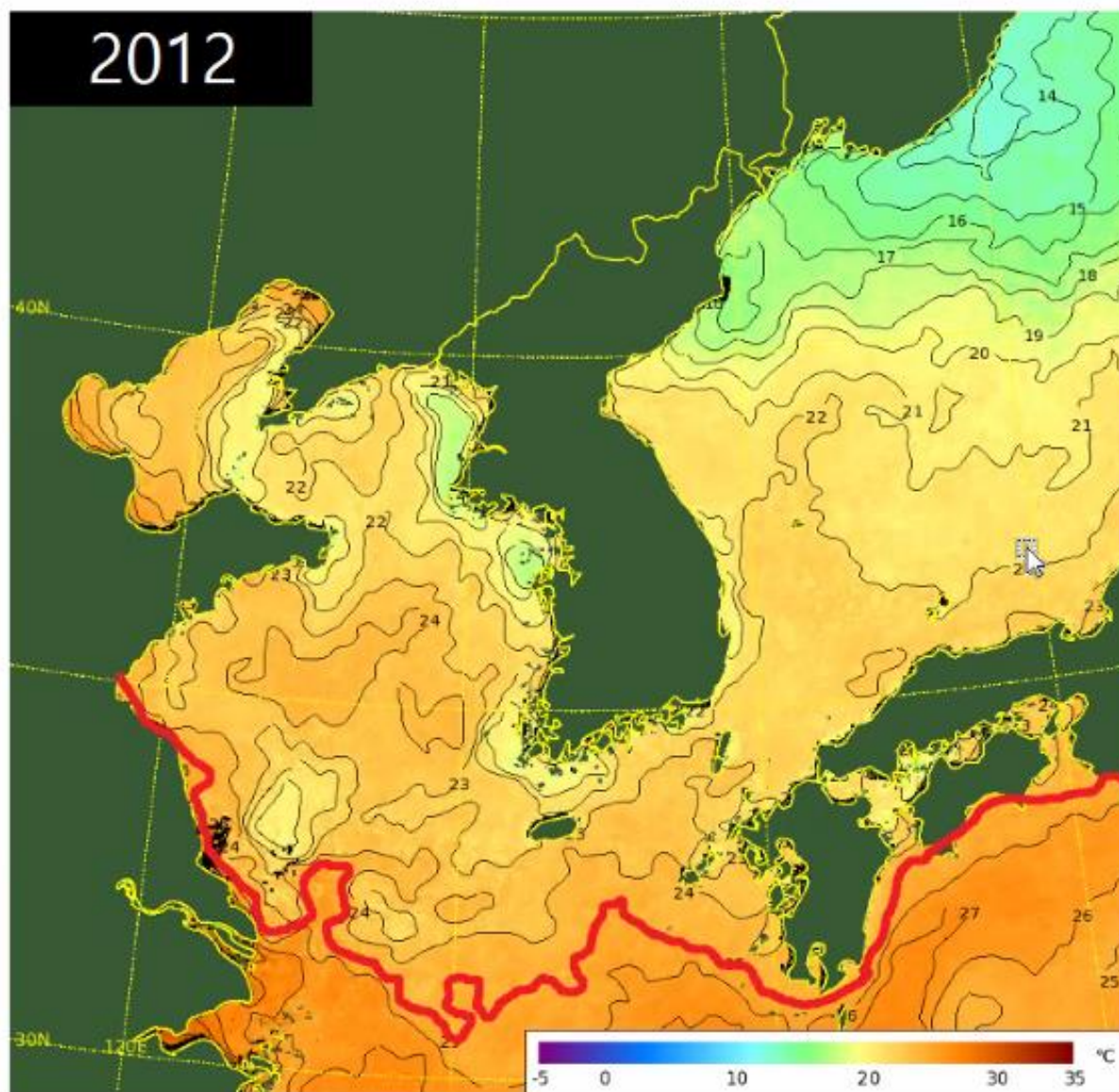
〈 해역별 1968-2019년 연평균 수온 추이 〉



출처\_국립수산물과학원 한국해양자료센터 정선해양관측자료

## 01 연구배경 및 필요성

〈 한반도 수온위성자료 (2012-2019 7월 평균) \_ 25°C 등수온선〉



## 01 연구배경 및 필요성



- 난류성 어종 증가 및 한류성 어종 감소 현상 극명화
- 기후 변화에 따른 어류, 패류 및 기타 생물 등 양식생산력 변동 및 양식적지 변화
- 새로운 아열대성 어종의 출현으로 신시장 개척

### 4만6천t 잡히던 명태, 작년엔 1t... 수온상승·남획에 '멸종위기'

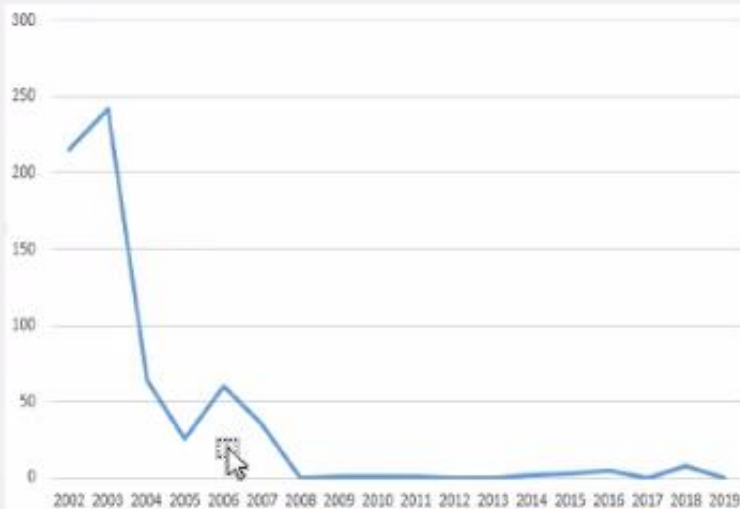
연합뉴스 2018.06.25

명태가 '멸종위기'에 가까워졌다.

명태 어획량은 1986년 4만6천890t에서 작년 1t으로 급감했다.

동해안 해역 수온이 상승하면서 명태가 북태평양으로 이동한데다가 어린 치어(노가리) 남획으로 자원량이 부족해져 2000년부터 어획량이 크게 줄었다고 통계청은 설명했다.

### < 명태 2002-2019년 어획량 변화 >



출처\_해양수산부 수산정보포털 어업생산통계 품종별현황

### 수온 1℃ 상승에... '멸치판'된 우리 바다 명태·꽁치는 자취 감춰

뉴데일리 2019.11.28

서해와 남해 할 것 없이 우리나라 연근해 전체를 멸치가 휩쓸고 있다.

2017년 기준으로 한 해 동안의 멸치 전체 어획량이 21만943톤이다. 서해·동해 가릴 것 없이 치고 올라오며 우리나라 연근해를 석권했다.

통계청은 "기후변화 영향으로 경남, 전남, 충남 등 전 해역으로 확대됐다"며 "연근해 해역에서 가장 많은 어획량을 차지한다"고 했다.



## 01 연구배경 및 필요성

☛ 수온 상승으로 인해 한류성 어종의 어획량 감소가 일어났다.

➡ 난류성 어종의 증가도 일으키지 않을까?

➡ 수온 상승으로 인한 아열대성 어종의 복상 ?



수온의 변화와 어종의 어획량의 변화는 관계성이 존재할 것이다.

### < 가설수립 >

$H_0$  VS  $H_1$

$H_0$  : 수온은 어종의 어획량에 관련이 없다.  $H_1$  : Not  $H_0$

- 수온과 어종의 어획량의 변화의 관계를 알아본다

### < 데이터 정규화 >

- 데이터의 값 크기 차이가 크다.
- 큰 값을 가지는 어획량의 단위를 kt (킬로톤) 으로 변경하여 수치의 크기를 정규화

### < 가설확인 p-value >

종속변수 : 해역별 어획량

유의수준 ( $\alpha$ ) : 0.05

해역	동해	유의 확률 (p)	8.72E-33
	남해		4.04E-34
	서해		3.45E-21

- 서식 해역별로 수온의 편차가 상이하여 해역 분리
- 해역별 서식 어종 또한 구분하여 총 어획량 계산

모든 해역 유의 확률이 0.05이하

귀무가설 기각 ➡ 수온은 어종의 어획량에 관련이 있다.

- 어획량은 수온 이외의 다양한 변수들에 영향을 받는 종속변수

➡ 수온 중심의 해양데이터 기반 수산자원 어획량 예측

## 02 데이터

### 2-1. Data Collection

수집데이터			건수	출처	비고
해양데이터 정선해양관측자료	수온	1968.01 - 2019.12	380,115	한국해양자료센터 국립수산과학원	해역별 분류
	염도		370,635		
	용존산소량		42,748		
기상데이터	연 평균 기온	1970 - 2019	50	기상청	
	태풍 발생 횟수	1970.01 - 2019.12	165	기상청	
추가적 해양데이터	클로로필	2004 - 2019	16개 년 / 160	해양환경공단	특정 어종 모델에서 2010 ~ 부터 사용
	화학적요구산소량				
	암모니아성 질소				
	아질산성 질소				
	질산성질소				
	용존무기질소				
	규산규소				
	총질소				
	총인				
	총용존무기인				
1인당 어업이익	1인당 어업이익	1970.01 - 2019.12	68	통계청	
어획량	어획량	1970.01 - 2019.12	18,758,968	수산정보포털 해양수산부	어종별 분류

## 02 데이터

### 2-2. 분석데이터 선정

#### < Data >

##### 해역별 해양데이터

1970 ~ 2019 년 시계열 데이터  
수온, 염도, 용존산소량 등

##### 외부측정데이터

기온, 태풍 등의 기상데이터 와  
추가적 해양데이터

##### 어종별 어획량

1970 ~ 2019 년 시계열 데이터  
어종별 상이

#### < Model >

어종변수



상관관계분석

ARIMA, fbprophet,  
Linear Regression

단변량 시계열 예측

LSTM, Xgboost,  
Multivariate Linear Regression

다변량 예측

어종 modeling

대표 어종군 선택

목표 변수 도출  
< 기대 어획량 >



## 02 데이터

### 2-3. 데이터 전처리

- 해역별 해양데이터, 외부측정데이터, 어획량 시계열 통합

1970 - 2019 년 시계열 데이터

- 목표변수가 되는 어획량 데이터가 1970~ 부터 나타남  
따라서 해양데이터와 외부측정데이터 또한 1970~ 로 통합

- 결측치 조정

보유 데이터셋의 비대칭성과, 해양데이터의 결측치 존재

- 측정일, 조사 기관의 다양함, 데이터 클래스 크기의 불균형  
해양데이터 특성상 측정장소에서 값 손실이 발생함

#### 해양데이터

결측치가 적은 월의 데이터 선택  
앞 뒤 데이터의 평균 값 기입 → 연 평균 값으로 조정

#### 외부측정데이터

기상데이터      기온      앞 뒤 데이터의 평균 값 기입 → 연 평균 값으로 조정

- 규칙적인 데이터 수 최대 확보

raw 해양 데이터의 경우 측정 일자, 시간대 등이 불규칙

- 측정일의 기상 상황 등에 따라 측정일의 변화가 생기고,  
2000년대 이전의 데이터는 수기작성으로 이상치가 나타남
- 측정 시간대의 불규칙으로 낮 과 밤의 온도차에 의해  
일,월 단위의 평균을 내기 어려움

해양 데이터의 여러 모델 구축 후 어획량데이터와 상관관계 분석  
가장 영향을 많이 미치는 시간대 선정 후 연도 별 평균 값 지정



새로운 변수 정의

year, 해역/월/낮or밤/수심	서/2/낮/0	서/4/낮/0	서/6/낮/0	서/8/낮/0	서/10/낮/C	서/12/낮/C
1970	5.579084	5.814175	11.60624	17.05638	17.85315	9.528984
1971	5.43731	6.213724	11.19929	17.01958	14.82353	10.60257
1972	6.838588	6.923267	11.52318	18.16311	17.14668	13.30064

⋮

## 2-3. 데이터 전처리

- 새로운 변수정의 - 통합할 수 있는 변수 와 나눌 수 있는 변수 설정

기존 변수	변수 설명	새로운 변수	변수 설명
통합할 수 있는 변수			
암모니아성질소+질산성질소 + ... +규산규소+총질소	같은 단위를 가지고 어류에 미치는 영향 이 비슷한 측정데이터	영양염류	기존 변수의 큰 카테고리 명
나눌 수 있는 변수			
수온	관측소 별 불규칙 측정데이터	해역 + 측정 월 + 측정시간대 + 측정 수심	어종의 어획에 알맞은 해역, 월 등의 수온 과 데이터의 규칙성을 맞춤

- 데이터의 코드화

- 변수를 코드화 함으로써 예측모델 적용에 용이하도록 함
- 해역별로 구분되는 변수는 변수 코드 앞에 각 해역명을 추가

어종명	F 0~8	클로로필	Chl-a
수온	Ocean_temp	화학적요구산소량	COD
염도	Salinity	영양염류	DOM
용존산소량	Oxy	1인 어업이익	Net_fishing
기온	Temp_avg	어획량	Fishing
태풍횟수	Typhoon		

- 확보하기 힘든 데이터 - 불법조업

- 데이터의 양이 적음

- 데이터화 된 내용은 최근 5년간 데이터만 가능

- 중국어선 나포 건수는 전체를 대변하기에 한계가 있음  
중국어선 입출역 데이터는 비공개 내부 데이터





## 03 분석 및 결과

### 3-1. 어종 Modeling

#### • 어종 modeling

	F0	F1	F2	F3	F4	F5	F6	F7
1	남/동해	서해	남/서해	남/동해	동해	동해	남/동해	남/동해
2	난류성	아열대성	난류성	아열대성	한류성	한류성	난류성	난류성
3	TAC O	TAC O	TAC X	TAC X	TAC X	TAC O	TAC O	TAC X
4	추세증가	추세증가	추세증가	추세증가	추세감소	추세감소	추세증가	추세증가

- TAC (Total Allowable Catch) : 총허용 어획량 제도
  - 개별어종(단일어종)에 대해 연간 잡을 수 있는 어획량을 설정하여 그 한도 내에서만 어획을 허용하여 자원을 관리하는 제도

☛ 어종이 가지는 개별적 특성과 어획량과 관련된 어종별 외부요인 변수 고려

동, 남, 서 해역 전반에 대해  
대표성을 지닌 8개의 대표 어종 선택

- 어종 모델의 TAC 유무의 경우  
인간 개입이 심한 어종 모델로 판단

- TAC로 관리되는 어종모델은  
불법조업, 치어남획 등 규칙성이 없고  
데이터 수집에 한계가 존재하여  
예측이 어려움

- 예시는 남해 서식 멸치모델로  
남해의 수온과 가장 높은 상관 관계  
를 나타냄

#### 멸치 F2

South\_Ocean\_temp

South\_Salinity

South\_Oxy

South\_DOM

South\_Ch1-a

South\_COD

Temp\_avg

Typhoon

Net\_fishing

해양  
데이터

해양  
추가  
데이터

기온  
데이터

Fishing

## 03 분석 및 결과

### 3-2. 상관관계 분석

#### 데이터 분석

- 어종 모델별 어획량을 종속변수로 가지는 상관관계분석 진행
- 피어슨의 상관계수 범위를 이용하여  $R^2 > 0.3$  이상을 강한 상관관계로 판단

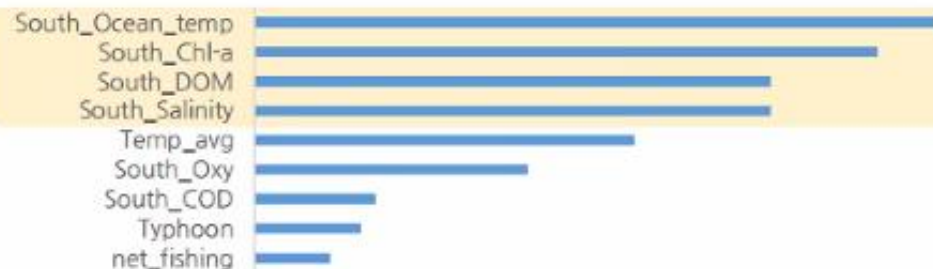
#### 분석모델 설계

- 종속 변수
  - Fishing\_total (어종별 해역 연간 어획량)
- 독립 변수
  - 해양데이터
    - Ocean\_temp (수온)
    - Salinity (염도)
    - Oxy (용존산소량)
  - 추가 해양데이터
    - DOM (영양염류)
    - Chl-a (클로로필)
    - COD (화학적 요구산소량)
  - 기후 데이터
    - Temp\_avg (기온)
    - Typhoon (태풍횟수)
  - Net\_fishing (1인당 어업이익)

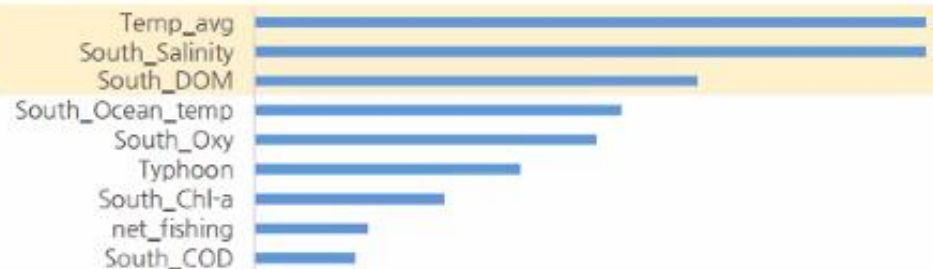
절대값으로 바꾸는게 좋음

#### 상관 관계 분석 결과

##### 멸치



##### 참다랑어



🐟 정확한 학습을 위해 상관관계가 큰 어종변수를 선택함

## 03 분석 및 결과

### 3-2. 단변량 시계열 예측

☛ 상관 관계가 높게 나온 변수에 대한 미래 데이터 단변량 예측

- ARIMA, fbprophet, Linear Regression 사용



#### Validation

- 각 어종 모델별 상관 관계가 높은 변수들을 예측모델에 적용
- 어종 모델별 해역별 데이터 모두 개별 적용
- 평균제곱오차(MSE)를 사용하여 가장 오차값이 낮은 모델 선정

선정된 모델의 단변량 예측 데이터를  
다변량 예측에 사용



## 03 분석 및 결과

### 3-3. 다변량 예측

☛ 단변량 예측으로 나온 미래 데이터로 다변량 예측

- LSTM, **XGBoost**, Multivariate Linear Regression 사용

☛ MinMaxScaler를 이용하여 데이터의 scale 조정

```
min_max_scaler = MinMaxScaler()
fitted = min_max_scaler.fit(data1)
print(fitted.data_max_)

data2 = min_max_scaler.transform(data1)
data2 = pd.DataFrame(data2, columns=data1.columns, index=list(data1.index.values))
data2
```

	South_Fishing	daytime_6_0	South_DOM	Temp_avg		South_Fishing	daytime_6_0	South_DOM	Temp_avg
0	290	17.170564	177.0	12.7	0	0.122731	0.686165	0.000000	0.307692
1	148	15.684702	508.0	12.4	1	0.000000	0.000000	0.231923	0.076923
2	241	16.838441	243.2	12.3	2	0.080380	0.532808	0.046385	0.000000
3	461	16.553626	542.1	12.9	3	0.270527	0.401278	0.255816	0.461538
4	1305	17.850097	1604.2	13.1	4	1.000000	1.000000	1.000000	0.615385
5	713	16.837825	574.9	13.4	5	0.488332	0.532523	0.278798	0.846154
6	912	17.319410	348.5	13.6	6	0.660328	0.754924	0.120165	1.000000
7	711	16.570190	645.0	13.1	7	0.486603	0.408927	0.327915	0.615385
8	539	17.493291	497.4	13.0	8	0.337943	0.835224	0.224496	0.538462
9	570	16.924238	313.6	13.5	9	0.364736	0.572430	0.095712	0.923077

```
XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
              importance_type='gain', interaction_constraints='',
              learning_rate=0.1, max_delta_step=0, max_depth=5,
              min_child_weight=1, missing=nan, monotone_constraints='()',
              n_estimators=100, n_jobs=0, num_parallel_tree=1,
              objective='reg:squarederror', random_state=0, reg_alpha=0,
              reg_lambda=1, scale_pos_weight=1, subsample=1, tree_method='exact',
              validate_parameters=1, verbosity=None)
```

learning\_rate : 0.1  
max\_depth : 5

```
B=model.predict(C)
B
```

```
array([[0.12166232, 0.00300843, 0.08074874, 0.1425643, 0.6527878,
        0.4867927, 0.5731224, 0.28828585, 0.3409496, 0.38295823,
        0.26536977, 0.2800886, 0.28558218, 0.45137778, 0.27130005,
        0.28101924, 0.5676288, 0.5676288, 0.28259543, 0.5686297,
        0.5686297, 0.5686297, 0.5686297, 0.5697221,
        0.5697221, 0.5686297, 0.5758076, 0.5758076, 0.5758076,
        0.57471514, 0.5727061, 0.5727061, 0.5727061, 0.5727061,
        0.5727061, 0.5727061, 0.6515795, 0.5727061, 0.5727061,
        0.6515795 ], dtype=float32)
```

#### Validation



- 0 과 1 사이의 값으로 정규화된 예측 결과를 원래의 값으로 디코딩하여 그래프에 나타냄
- 평균제곱오차(MSE) 0.0706

## 03 분석 및 결과

### 3-3. 다변량 예측

단변량 예측으로 나온 미래 데이터로 다변량 예측

- LSTM, XGBoost, Multivariate Linear Regression 사용

MinMaxScaler를 이용하여 데이터의 scale 조정

```
min_max_scaler = MinMaxScaler()
fitted = min_max_scaler.fit(day1)
print(fitted.data_max_)
day2 = min_max_scaler.transform(day1)
day2 = pd.DataFrame(day2, columns=day1.columns, index=list(day1.index.values))
day2
```

	South_Fishing	daytime_6_0	South_DOM	Temp_avg
0	290.0	17.170564	177.000000	12.700000
1	148.0	15.684702	508.000000	12.400000
2	241.0	16.838441	243.200000	12.300000
3	461.0	16.553626	542.100000	12.900000
4	1305.0	17.850097	1604.200000	13.100000
5	713.0	16.837825	574.900000	13.400000
6	912.0	17.319410	348.500000	13.600000
7	711.0	16.570190	645.000000	13.100000
8	539.0	17.493291	497.400000	13.000000
9	570.0	16.968056	313.600000	13.500000
10	NaN	16.976273	186.326844	13.032612
11	NaN	17.031534	250.869972	13.208156
12	NaN	17.056560	311.875380	13.317979



	South_Fishing	daytime_6_0	South_DOM	Temp_avg
0	0.122731	0.686185	0.075095	0.224310
1	0.000000	0.000000	0.289602	0.056077
2	0.080380	0.532808	0.117997	0.000000
3	0.270527	0.401278	0.311700	0.336464
4	1.000000	1.000000	1.000000	0.448619
5	0.488332	0.532523	0.332957	0.616851
6	0.660328	0.754924	0.186237	0.729006
7	0.486603	0.408927	0.378385	0.448619
8	0.337943	0.835224	0.282732	0.392542
9	0.364736	0.592665	0.163620	0.672929
10	NaN	0.596460	0.081140	0.410942
11	NaN	0.621980	0.122967	0.509270
12	NaN	0.633537	0.162502	0.570856

Epoch : 1000      Batch size : 3

Layer (type)	Output Shape	Param #
lstm_4 (LSTM)	multiple	4608
dense_4 (Dense)	multiple	33

Total params: 4,641  
Trainable params: 4,641  
Non-trainable params: 0

```
pred2 = net.predict(test_feature2)
pred2
```

```
array([[0.34733087],
       [0.14929172],
       [0.26913232],
       [0.32957035],
       [0.60563815],
       [0.43540764],
       [0.48906046],
       [0.35133642],
       [0.45210272],
       [0.42854536],
       [0.36163384],
       [0.39664936],
       [0.41820067],
       [0.4272918 ],
       [0.4059351 ],
       [0.4077446 ],
       [0.42896295],
       [0.43729484],
       [0.4299565 ],
       [0.45766932],
       [0.473126 ],
       [0.47713685],
       [0.45402247],
       [0.4818717 ],
       [0.49628866],
       [0.50818674],
       [0.47763556],
       [0.5043038 ],
       [0.5192822 ],
       [0.5230775 ],
       [0.5010655 ],
       [0.52735746],
       [0.5421198 ],
       [0.5458219 ],
       [0.52430904],
       [0.5502465 ],
       [0.5648115 ],
       [0.56855416],
       [0.5473658 ],
       [0.57298213]], dtype=float32)
```

Validation

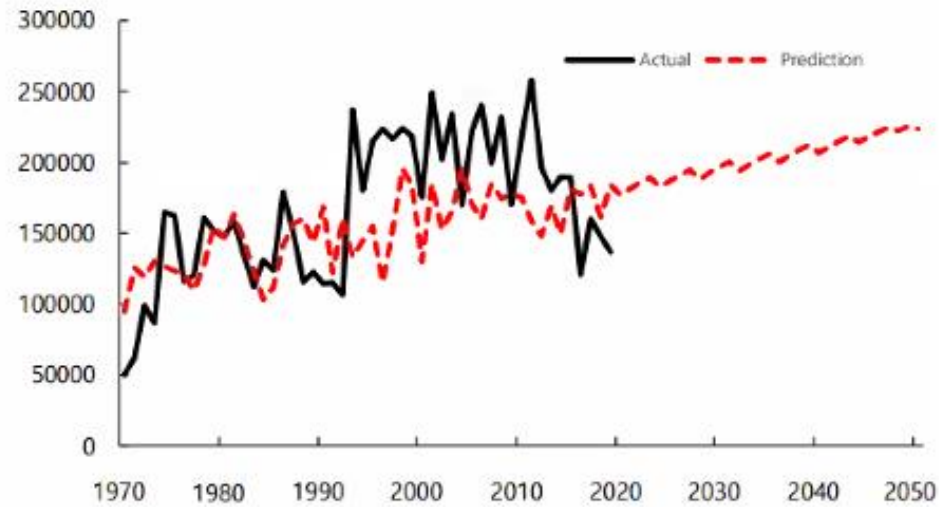


평균제곱오차(MSE) 0.0472  
가장 낮은 값을 나타냄

## 04 적용과 기대효과

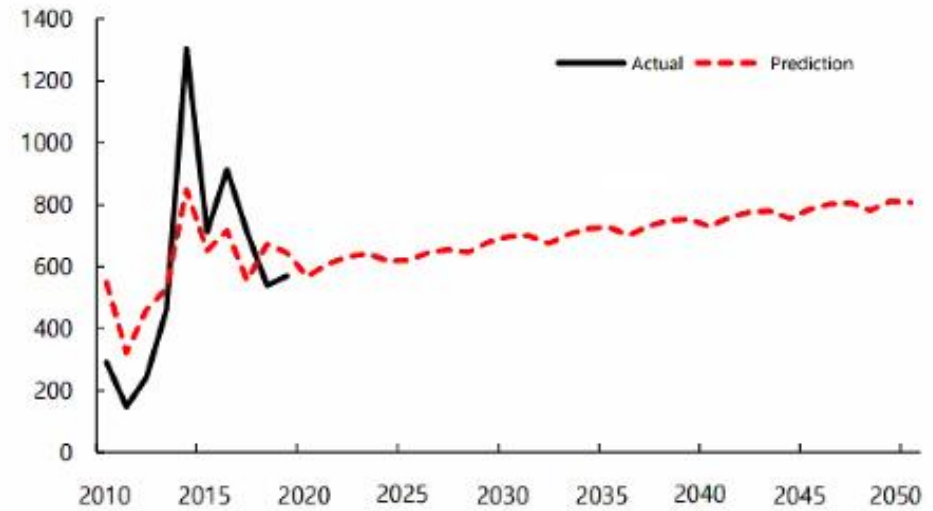
### 4-1. 실제적용

남해\_멸치\_예측그래프



- 어획량이 집계된 1970년 기준 어획량 데이터를 사용하여 2050년까지 향후 30년 어획량 예측
- MSE(평균제곱오차) 값 : 0.047

남해\_참다랑어\_예측그래프



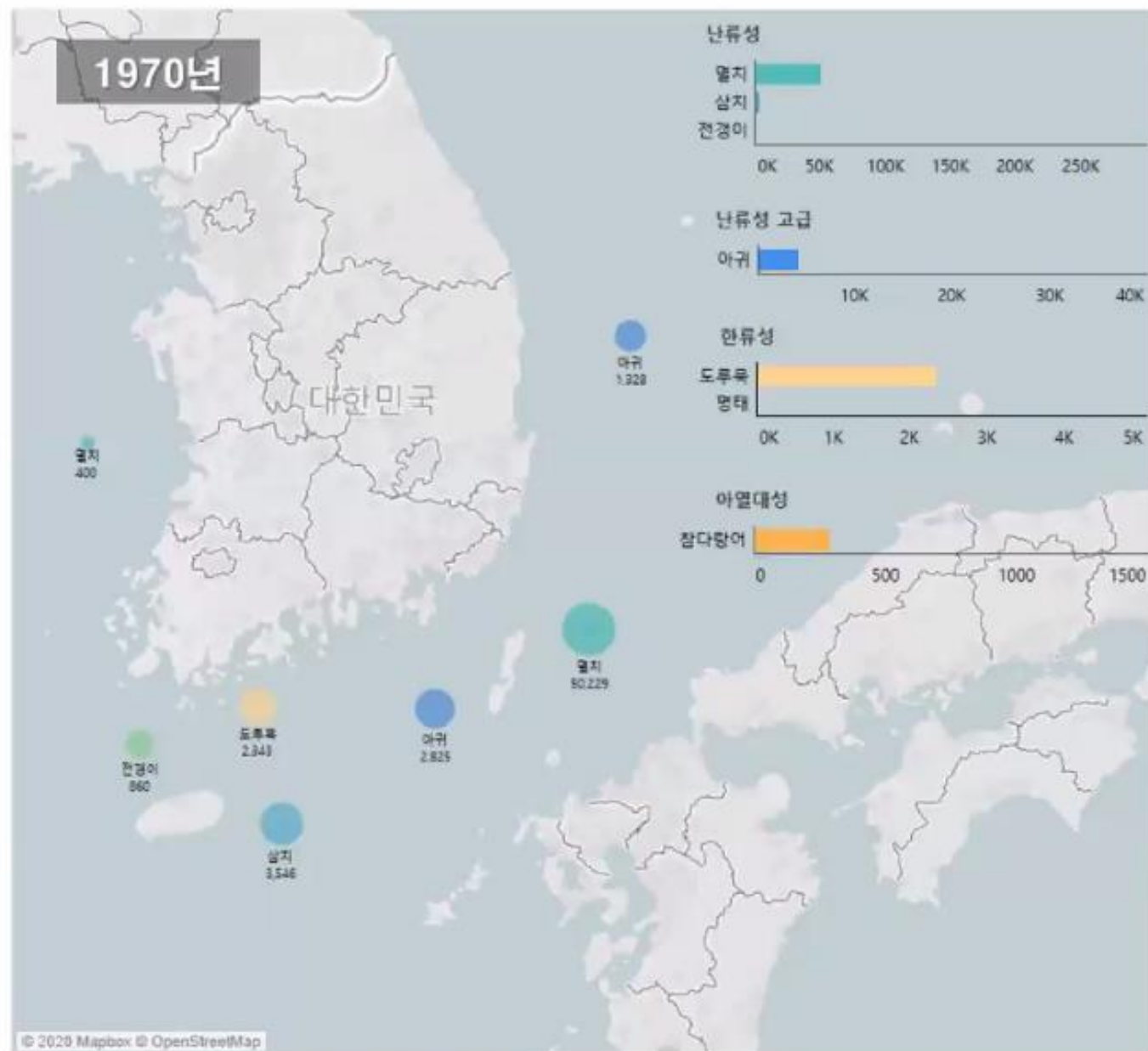
- 어획량이 집계된 2010년 기준 어획량 데이터를 사용하여 2050년까지 향후 30년 어획량 예측
- MSE (평균제곱오차) 값 : 0.068



## 4-2. Fish Farm Road

어획량 예측지도

물고기 크기별로  
그리는데  
조심



## 04 적용과 기대효과

### 4-4. 기대효과

- 미래 생산성이 높아지는 어종을 미리 파악하여 선제적 대응
- 어획 및 양식에 필요한 장비 혹은 시설을 미리 구비할 수 있음



- 다양한 어종 수급 정보를 활용하여 원산지를 파악해 스마트한 소비 가능
- 소비자 수산물 가격 안정

- 국내산 어종 수급의 변화에 대한 발 빠른 대처
- 수입에 의존하고 있었던 인기 어종들의 국내 시장 개척

모델 정확도 향상 & 확장 → 어장에 직접적인 피해를 미치는 새로운 아열대성 해파리 추적으로 어민 피해 최소화



## 05 참고 문헌

- 엄기혁 외, 2015, 기후변화에 따른 한국 연근해 어업생산량 변화 분석, 한국수산경영학회 31-41
- 김종춘, 2011, 기후변화 현상과 국제 동향, 오토 저널 68-74
- T. Wernberg *et al*, 2011, Impacts of climate change in a global hotspot for temperate marine biodiversity and ocean warming, Science Direct Volume 400
- Jorge Garcia Molinos *et al*, 2016, Climate velocity and the future global redistribution of marine biodiversity, nature climate change 83-88
- Grégory Beaugrand *et al*, 2010, Marine biodiversity, ecosystem functioning, and carbon cycles, PNAS 10120-10124
- Francisco Ramírez *et al*, 2017, Climate impacts on global hot spots of marine biodiversity, ScienceAdvances Vol. 3
- 유준택 외, 2018, 가을철 남해안의 멸치 어황 예측방법, 대한민국 특허청 10-1836813
- 유신재, 장창익, 1993, Forecasting of Hairtail (*Trichiurus lepturus*) Landings in Korean Waters by Times Series Analysis, Journal of the Korean Fisheries Society v.26 363-368
- 이승종 외, 2011, Long-term variation in catch of spanish mackerel (*Scomberomorus niphonius*) related to environmental change in Korean waters, 한국수산해양기술학회 47 (2) : 99-107
- 조용준 외, 2006, 일반해면어업 어획량의 시계열 분석, 농촌경제 29(1) 123-134
- 김준택 외, 2002, Characteristics of Catch Fluctuation and Distribution of Yellow tail, *Seriola quinqueradiata*, TEMMINCK et SCHLEGEL, in Korean Waters, 한국어업기술학회지 v.38 no.1, 11-19
- 조정희 외, 2009, 생물경제모형을 이용한 수산물 최적 생산량 추정 및 활용에 관한 연구, 한국해양수산개발원, 2009-14
- 곽나영, 2020, 수온에 따른 어획량 예측과 어류 및 수산 종목 소비자물가지수의 개입, 통계연구 21 (0): 87-100