

접수번호	※ 해당 없음
------	---------

「제5회 해양수산 공공데이터 및 빅데이터 활용 경진대회」

기 획 서 (빅데이터 아이디어 기획 및 서비스 부문)

\* 해당란에 ㉠ 표시

1. 참가자 정보		
개인·팀·기업명	JellyFish Under The Sea.	
서비스(제품)명	해양 환경에 따른 해파리 출몰 분석	
공모 분야	* <input type="checkbox"/> 빅데이터 아이디어 기획 및 서비스 부문	* <input type="checkbox"/> ① 해양 <input type="checkbox"/> ② 수산 <input type="checkbox"/> ③ 해운물류 <input type="checkbox"/> ④ 해사안전 <input type="checkbox"/> ⑤ 항만
서비스 유형	<input type="checkbox"/> 모바일(태블릿pc) 앱·웹 서비스 * <input type="checkbox"/> 온라인 웹 서비스 <input type="checkbox"/> 웨어러블 디바이스	<input type="checkbox"/> 제품제작 <input type="checkbox"/> 영상 및 시각화 <input type="checkbox"/> 기타 (                      )

2. 기획서 작성
1. 개요
1.1) 아이디어 기획 핵심내용, 제안배경(구체성, 우수성)
<p><b>가. 개요</b></p> <p>최근 국내 연안에 급증한 해파리로 인하여 피해사례가 끊이지 않고 있다. 이를 해결하기 위해 정부에서도 전국 단위의 중장기 계획을 수립하고 많은 시간과 비용을 들이고 있지만 큰 성과를 나타내지 못하고 있는 실정이다. 이를 해결하기 위해 빅데이터 분석을 활용하여 해양 환경 정보로 해파리의 출몰 확률을 도출하려한다.</p>

## 나. 개선되지 않는 해파리 피해



최근 해안의 수온이 상승함에 따라 해파리 개체 수가 증가하면서 어업에서 2290억원, 발전소에서 588억원, 그리고 관광 분야에서 170억원 등 그 경제적 피해액이 3000억에 달하고 있다. 그 예시로 위의 사진처럼 어업활동에 물고기보다는 해파리가 많음을 볼 수 있다. 그러나 해파리 퇴치는 고비용, 저효율의 수작업으로 많은 어려움이 따른다. 이에 2014년, 해양수산부와 미래창조과학부는 해파리 통합방제시스템을 구축하여 해결하려 했지만 큰 성과를 거두지 못했다. ‘스마트부이’는 해파리의 대량 출현 및 해양 정보를 탐지하는 부표형태의 기구로 해양수산부와 미래창조과학부의 주관 하에 만들어졌다. 스마트부이는 해파리의 대량 출현 탐지 및 이동경로를 예측하고, 이 위치정보를 지능형 퇴치 로봇에게 전송하면 그 퇴치 로봇이 GPS 기반 자율주행 기능으로 해파리를 추적, 분쇄하여 제거하는 형식이다. 이 시스템으로 해파리 제거에 소요되는 과도한 인력 및 시간을 경감하고 신속한 대응으로 국민 피해를 최소화시킬 것이라는 기대를 모았지만 그 효과는 크지 않았다. 이에 해양 환경 정보를 통해 해파리의 출몰을 예측하는 아이디어를 제안하게 되었다.

## 다. 아이디어 세부내용

우선 국립수산물과학원의 해파리 출몰 데이터와 어장환경모니터링 데이터를 수집한다. 두 데이터를 같은 기간과 장소로 결합하기 위해 해파리 출몰 지역의 위도 경도 값을 구글 지도를 통해서 도출하고 이 위도, 경도에 따라 가장 가까운 어장환경모니터링 정점을 파악하여 두 데이터를 결합한다. 이때, 해파리 출몰을 0(출몰하지 않음), 1(출몰함)의 범주형으로 데이터화한다. 이를 바탕으로 같은 기간과 장소의 해양 환경의 변수(수온, 염분, 용존산소량 등)를 독립변수로 두고 범주형 데이터인 해파리 출몰을 종속변수로 로지스틱 회귀분석을 실시한다. 분석을 통해 해파리 종류별로 출몰에 영향을 받는 해양 환경 변수를 도출할 수 있다. 영향도가 높은 해양 환경 변수로 해파리가 출몰할 확률을 도출한다. 학습된 로지스틱 회귀 모형으로 해양

환경을 조사하는 것만으로 해파리 출몰 여부를 예측하고 피해에 대한 대비가 가능해진다.

자료 출처:

해양수산과 수산행정 홈페이지

“창조경제 대표사례, 로봇으로 해파리 퇴치(수산자원정책과)”, 2014.10.01

강진우,진은석, MBC 뉴스 기사,” ‘물 반 해파리 반’... 속 터지는 어민들”, 2017.08.10

(<http://www.mbn.co.kr/pages/vod/programView.mbn?bcastSeqNo=1162156>)

김준래, “해파리 퇴치 위한 장기계획 나왔다”, The Science Times, 2018.02.23

## 1.2) 활용데이터(해양수산분야)

시스템 : 국가해양환경정보통합시스템(<http://meis.go.kr/rest/main>) - 국립수산물과학원

데이터명 : 해파리 출몰 정보, 어장 환경 모니터링

해파리 출몰 정보

jellyFishName	newsPerSt	newsPerEn	density	spot
노무라입깃해파리	20141003	20141010	소량	인천 소청도
노무라입깃해파리	20141003	20141010	소량	인천 연평도
노무라입깃해파리	20141003	20141010	소량	충남 격렬비열도
노무라입깃해파리	20141003	20141010	소량	전북 군산 개야도

Open API를 활용하여 해파리의 이름, 조사 시작 날짜, 조사 종료 날짜, 출몰한 규모, 출몰 지역의 데이터를 만듦

jellyFishName	노무라입깃해파리	보름달물해파리	커튼원양해파리	라스톤입방해파리	미확인	Aurelia limbata	야광원양해파리	오이빛해파리
newsPerSta	20141003	20141011	20141017	20141024	20151002	20151009	20151204	20160610
newsPerEnd	20141010	20141016	20141023	20141030	20151008	20151015	20151210	20160616
density	소량	밀집						
spot	인천 소청도	인천 연평도	충남 격렬비열도	전북 군산 개야도	전북 고창	전북 왕등도	전남 진도	전남 완도

해파리 종류 : 총 7종 (미확인 제외)

조사 기간 : 2014년 10월 부터 2016년 6월

출몰한 규모 : 소량/밀집

출몰한 장소 : 총 155개 지역

어장 환경 모니터링

해역명	연안명	정점명	관측년도	관측월	관측일	tude [degrees]	ide [degrees]	수온(°C)		염분(psu)	
								표층	저층	표층	저층
서해중부(VII)	가로림연안	가로림 01	2016	6	11	126.311111	36.9930556	14.17	14.15	32.03	32.03
서해중부(VII)	가로림연안	가로림 01	2015	12	12	126.311111	36.9930556	11.65	11.42	31.5	31.49
서해중부(VII)	가로림연안	가로림 01	2015	10	8	126.311111	36.9930556	21.53	21.36	31.82	31.9
서해중부(VII)	가로림연안	가로림 01	2015	8	8	126.311111	36.9930556	21.45	24.98	31.77	31.77
서해중부(VII)	가로림연안	가로림 01	2015	6	5	126.311111	36.9930556	13.13	13.14	31.84	31.84

#### 국립수산과학원에 어장 환경 모니터링을 다운로드

해역명	서해중부(VII)	남해서부(V)	동해남부(II)	동해중부(I)	제주도(VIII)	남해중부(IV)	남해동부(III)	서해남부(VI)
연안명	가로림연안	가막만	감포연안	강구연안	강릉연안	서귀포연안	거제도남안	거제도동안
정점명	가로림 01	가로림 02	가로림 03	가로림 04	가로림만 01	가로림만 02	가로림만 03	가로림만 04

해역명 : 총 8개 해역

연안명 : 총 56개 연안

정점명 : 총 406개의 정점

관측날짜 : 관측년도, 관측월, 관측일

위도/경도 : 각 지역의 위도와 경도

해양 성분 : 총 16가지

수온(°C)	염분(psu)	pH
용존산소량(mg/L)	화학산소요구량(mg/L)	암모니아성질소(μg/L)
아질산성질소(μg/L)	질산성질소(μg/L)	용존무기질소(μg/L)
총질소(μg/L)	용존무기인(μg/L)	총인(μg/L)
규산규소(μg/L)	부유물질(mg/L)	클로로필a(μg/L)
투명도(m)		

#### 1.3) 기존 서비스와의 차별성 및 독창성(창의성, 차별성)

기존의 해양수산부와 미래창조과학부에서 제작한 해파리 모니터링 기구 ‘스마트부이’는 해파리의 대량 출현 및 해양 정보를 탐지하는 부표형태의 기구로 카메라 등을 활용하여 해파리 군집을 탐색한다. 우리가 제안하는 분석기술은 일반적인 모니터링 기술이 아닌 해파리가 서식하는 환경을 분석하여 예측한다는 차이점이 있다. 기존의 방식은 해파리들이 카메라에 감지가 되어야 그 위치정보를 알 수 있었기 때문에 범위가 넓은 해양에서 사용하는 데에는 한계가 있었다. 그러나 해파리 출몰 분석을 활용하면 해양환경 성분에 따른 해파리의 출몰지점을 미리 예측할 수 있어 더욱더 효율적으로 해파리를 탐색할 수 있다. 해파리 서식환경의 데이터는 수온, 염분, pH, 용존산소량 등 16가지의 성분들을 토대로 정밀하게 분석되며, 이를 활용한 해파리 발생확률을 퍼센트로 제공하여 보다 더 세밀한 정보제공이 가능하다.

## 2. 데이터 분석 방법 및 시각화(분석의 적절성)

### 2.1) 분석방법(기법)

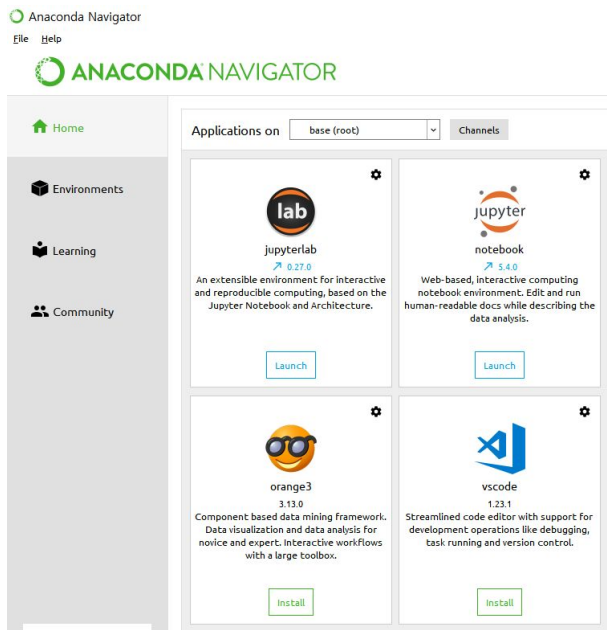
- 활용데이터를 이용한 분석 방법(기법) 작성

※ 분석방법 및 알고리즘 등 작성

※ 샘플데이터 요구시 데이터 제공 가능(단 개인정보, 보안정보 제외)

#### 분석 도구

anaconda jupyter notebook에서 python을 활용



#### 데이터 정제 및 전처리

분석 정제 및 전처리는 python의 패키지 pandas의 Dataframe 구조를 활용였다.

어장 환경 모니터링 데이터와 해파리 출몰 데이터에서 날짜와 지역이 같은 행을 한 행으로 묶고, 해파리가 출몰한 여부를 0(출몰하지 않음), 1(출몰함)으로 변환하여 분석 데이터 셋을 구축했다.

- 최종 데이터 셋의 열의 정보

해양정보	해역명	연안명	정점명	출현일	longitude	latitude
	수온(°C)	염분(psu)	pH	용존산소량(mg/L)	화학산소요구량(mg/L)	암모니아성질소( $\mu\text{g/L}$ )
해양성분	아질산성질소( $\mu\text{g/L}$ )	질산성질소( $\mu\text{g/L}$ )	용존무기질소( $\mu\text{g/L}$ )	총질소( $\mu\text{g/L}$ )	용존무기인( $\mu\text{g/L}$ )	총인( $\mu\text{g/L}$ )
	규산규소( $\mu\text{g/L}$ )	부유물질(mg/L)	클로로필a( $\mu\text{g/L}$ )	투명도(m)		
해파리	노무라입깃해파리	보름달물해파리	커튼원양해파리	라스톤입방해파리	Aurelia limbata	
	야광원양해파리	오이빛해파리	미확인			

#### 분석 방법 및 알고리즘

분석 방법은 python의 패키지 sklearn과 statsmodels를 활용하여 로지스틱 회귀분석을 실시하였다.

- 데이터 셋을 Train Data와 Test Data로 나누어 Train Data로 로지스틱 회귀분석 모델을 만들고 Test Data로 예측을 하여 모델의 성능을 평가했다.

## 2.2) 분석결과 및 시각화

### - 2.1) 분석 후 결과 값 및 결과를 활용한 시각화 내용(방법) 작성

#### 해양 환경 정보와 노무라입깃해파리의 로지스틱 회귀분석 결과

Logit Regression Results						
Dep. Variable:	보름달물해파리		No. Observations:	728		
Model:	Logit		Df Residuals:	711		
Method:	MLE		Df Model:	16		
Date:	Wed, 30 May 2018		Pseudo R-squ.:	0.08535		
Time:	17:54:12		Log-Likelihood:	-132.33		
converged:	True		LL-Null:	-146.28		
			LLR p-value:	0.08255		
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-0.6883	10.655	-0.065	0.948	-21.571	20.194
temp	-0.0463	0.074	-0.625	0.532	-0.192	0.099
salt	-0.2586	0.178	-1.449	0.147	-0.608	0.091
pH	0.5293	1.348	0.393	0.695	-2.113	3.172
DO	0.5068	0.184	2.751	0.006	0.146	0.868
COD	-0.4815	0.355	-1.357	0.175	-1.177	0.214
NH3N	-35.4866	27.749	-1.279	0.201	-89.873	18.900
NH2N	-45.0220	28.081	-1.603	0.109	-100.060	10.017
NO3N	-31.2242	32.477	-0.961	0.336	-94.878	32.430
DIN	39.0793	27.606	1.416	0.157	-15.027	93.185
TN	-1.0836	2.516	-0.431	0.667	-6.016	3.848
DIP	-66.0893	29.184	-2.265	0.024	-123.289	-8.889
TP	6.0866	8.459	0.720	0.472	-10.492	22.665
SiO2	2.6908	1.073	2.509	0.012	0.588	4.793
SS	0.0084	0.006	1.384	0.166	-0.003	0.020
BRICa	-0.0264	0.027	-0.962	0.336	-0.080	0.027
tansparency	-0.1120	0.073	-1.527	0.127	-0.256	0.032

p-value가 0.05보다 작은 해양 환경변수는 HN3N(암모니아성질소), NH2N(아질산성질소), DIN (용존무기질소), Tansparency(투명도)로 노무라입깃해파리의 출몰에 영향을 주는 변수이다.

각각의 coef를 확인해보면 HN3N과 NH2N, Tansparency의 수치가 높아지면 노무라입깃해파리가 출몰할 확률이 낮아지는 음의 관계와 DIN의 수치가 높아지면 출몰할 확률이 높아지는 양의 관계를 도출했다.

유의미한 변수로 다시 로지스틱 회귀분석을 실시



Logit Regression Results						
Dep. Variable:	노무라입깃해파리		No. Observations:	728		
Model:	Logit		Df Residuals:	723		
Method:	MLE		Df Model:	4		
Date:	Thu, 31 May 2018		Pseudo R-squ.:	0.08542		
Time:	11:12:52		Log-Likelihood:	-111.47		
converged:	True		LL-Null:	-121.88		
			LLR p-value:	0.0003433		
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-2.2456	0.447	-5.022	0.000	-3.122	-1.369
NH3N	-28.5020	10.897	-2.615	0.009	-49.861	-7.144
NH2N	-28.5451	10.922	-2.613	0.009	-49.953	-7.138
DIN	25.6581	9.880	2.597	0.009	6.294	45.023
tansparency	-0.3229	0.113	-2.861	0.004	-0.544	-0.102

NH3N이 한 단위 증가할 경우, 노무라입깃해파리 출몰할 확률은  $\exp(-28.5020)$ , 약 100감소한다.

NH2N이 한 단위 증가할 경우, 노무라입깃해파리 출몰할 확률은  $\exp(-28.5451)$ , 약 100배 감소한다.

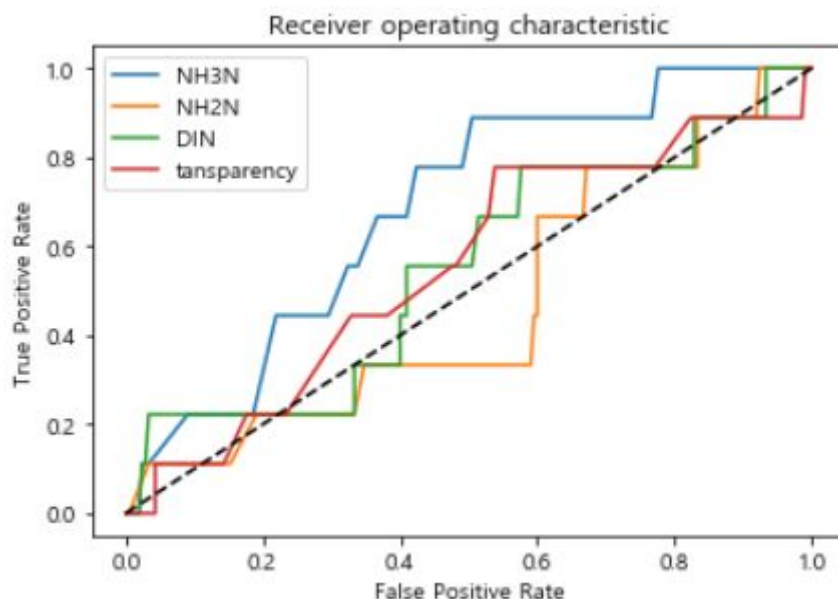
Tansparency이 한 단위 증가할 경우, 노무라입깃해파리 출몰할 확률은  $\exp(-0.3229)$ , 약 27배 감소한다.

DIN이 한 단위 증가할 경우, 노무라입깃해파리 출몰할 확률은  $\exp(25.6581)$ , 약 139050080979 배 증가한다.

노무라입깃해파리 출현 확률을 구하는 함수식

$$\frac{1}{1 + e^{-2.25 - 28.5 \cdot \text{NH3N} - 28.55 \cdot \text{NH2N} + 25.6581 \cdot \text{DIN} - 0.32 \cdot \text{tansparency}}}$$

학습된 모델의 노무라입깃해파리 출몰 예측 평가 - ROC 그래프



NH3N과 transparency는 노무라입깃해파리의 출몰 여부를 올바르게 구분해낼 확률이 옳바르지 못하게 구분해낼 확률보다 크다는 점을 확인할 수 있다. 하지만 NH2N와 DIN는 그래프로 확인 할 수 있듯이 노무라입깃해파리의 출몰을 옳바르게 예측하지 못한다는 것을 할 수 있다.

즉 위의 분석 모델에서 NH3N과 transparency는 노무라입깃 해파리를 출몰 여부를 파악하는데 유의미한 변수이다.

#### 해양 환경 정보와 보름달물해파리의 로지스틱 회귀분석 결과

Logit Regression Results						
Dep. Variable:	보름달물해파리	No. Observations:	728			
Model:	Logit	Df Residuals:	711			
Method:	MLE	Df Model:	16			
Date:	Thu, 31 May 2018	Pseudo R-squ.:	0.08535			
Time:	11:02:08	Log-Likelihood:	-132.33			
converged:	True	LL-Null:	-146.28			
		LLR p-value:	0.08255			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-0.6883	10.655	-0.065	0.948	-21.571	20.194
temp	-0.0463	0.074	-0.625	0.532	-0.192	0.099
salt	-0.2586	0.178	-1.449	0.147	-0.608	0.091
pH	0.5293	1.348	0.393	0.695	-2.113	3.172
DO	0.5068	0.184	2.751	0.006	0.146	0.868
COD	-0.4815	0.355	-1.357	0.175	-1.177	0.214
NH3N	-35.4866	27.749	-1.279	0.201	-89.873	18.900
NH2N	-45.0220	28.081	-1.603	0.109	-100.060	10.017
NO3N	-31.2242	32.477	-0.961	0.336	-94.878	32.430
DIN	39.0793	27.606	1.416	0.157	-15.027	93.185
TN	-1.0836	2.516	-0.431	0.667	-6.016	3.848
DIP	-66.0893	29.184	-2.265	0.024	-123.289	-8.889
TP	6.0866	8.459	0.720	0.472	-10.492	22.665
SiO2	2.6908	1.073	2.509	0.012	0.588	4.793
SS	0.0084	0.006	1.384	0.166	-0.008	0.020
BRICa	-0.0264	0.027	-0.962	0.336	-0.080	0.027
transparency	-0.1120	0.073	-1.527	0.127	-0.256	0.032

p-value가 0.05보다 작은 해양 환경변수는 DO(용존산소량), DIP(용존무기인), SiO2(규산규소), 로 보름달물해파리의 출몰에 영향을 주는 변수이다. 각각의 coef를 확인해보면 DIP의 수치가 높아지면 보름달물해파리가 출몰할 확률이 낮아지는 음의 관계와 DO와 SiO2의 수치가 높아지면 출몰할 확률이 높아지는 양의 관계를 도출했다.

유의미한 변수로 다시 로지스틱 회귀분석을 실시



Logit Regression Results						
=====						
Dep. Variable:	보름달물해파리		No. Observations:		728	
Model:	Logit	Df Residuals:	724			
Method:	MLE	Df Model:	3			
Date:	Thu, 31 May 2018	Pseudo R-squ.:	0.08589			
Time:	09:47:15	Log-Likelihood:	-141.03			
converged:	True	LL-Null:	-146.28			
		LLR p-value:	0.01476			
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	-6.4453	1.221	-5.278	0.000	-8.839	-4.052
DO	0.4437	0.145	3.058	0.002	0.159	0.728
DIP	-27.9823	18.925	-1.479	0.139	-65.075	9.111
SiO2	1.1440	0.594	1.925	0.054	-0.021	2.309
-----						

DO가 한 단위 증가할 경우, 보름달물해파리가 출몰할 확률은  $\exp(0.4437)$ , 약 156 배 증가한다.

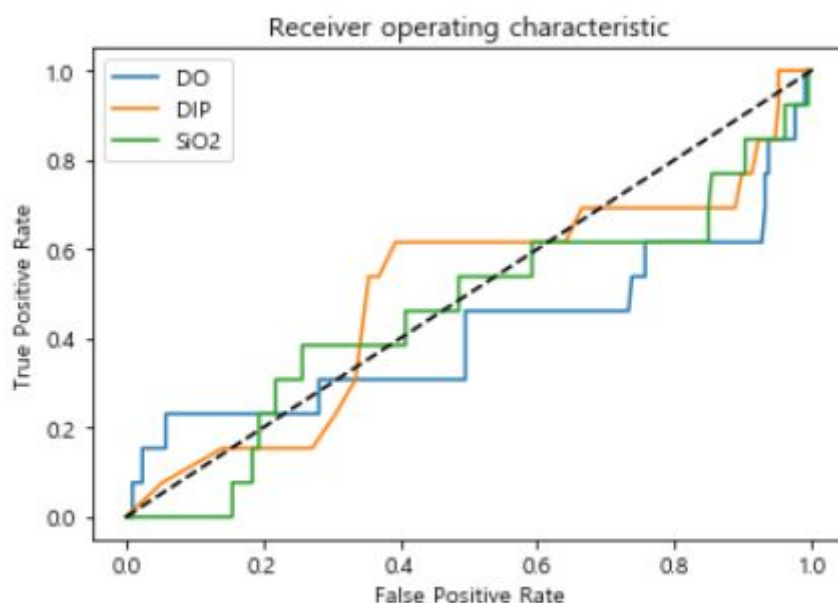
DIP가 한 단위 증가할 경우, 보름달물해파리가 출몰할 확률은  $\exp(-27.9823)$ , 약 100 배 감소한다.

SiO2가 한 단위 증가할 경우, 보름달물해파리가 출몰할 확률은  $\exp(1.1440)$ , 약 314 배 증가한다.

보름달물해파리 출현 확률을 구하는 함수식

$$\frac{1}{1 + e^{-6.45 + 0.44*DO - 27.98*DIP + 1.14*SiO2}}$$

학습된 모델의 보름달물해파리 출몰 예측 평가 - ROC 그래프



학습된 모델에서 DO, DIP, SiO2가 보름달물해파리 출몰에 영향을 준다는 결과가 나왔지만 위의

그래프에서는 모델의 성능이 좋지않다는 결과를 얻을 수 있었다.

#### 결론

기존 어장 환경 모니터링 데이터를 활용하여 피해를 주고있는 해파리의 출몰을 예측함으로써 데이터를 수집하는 데 드는 추가적인 비용없이 해파리 출몰에 대응하고 해파리 연구에 보탬이 될 수 있다.

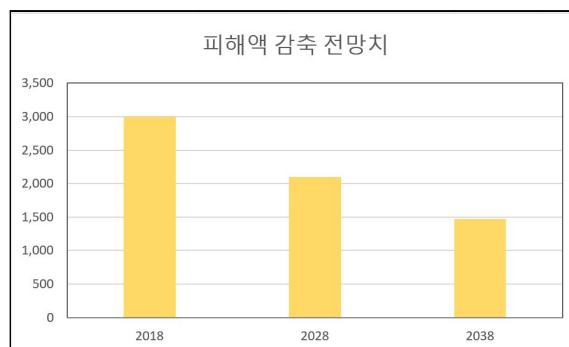
### 3. 기대효과 및 사업화

#### 3.1) 아이디어의 시장성 및 실현가능성(사업성, 실현가능성)

##### 가. 아이디어의 시장성 및 사업화 가능성

현재 스마트부이는 해파리의 출현상황을 모니터링 하여 기지국에 보내는 알고리즘만 존재한다. 하지만 해양환경에 따른 해파리 출몰 분석의 알고리즘을 추가한다면 해양의 정보에 따라 해파리가 출현할 확률을 알 수 있어 어업 종사자 및 피서객의 피해를 줄일 수 있다. 최근 급격하게 증가하고 있는 해파리로 인해 발생하던 문제점이 해결될 가능성이 높아지기 때문에 시장성이 충분히 높다고 생각한다. 더 나아가 해파리 출몰정보를 웹페이지에 제공한다면 어업종사자 및 해양관련자들에게 많은 도움이 될 것이기 때문에 사업가능성 또한 높다고 생각한다.

##### 나. 실현가능성 및 피해액감축전망



기존에 존재하는 스마트부이에 확률변수까지 추가하여 아직 출현은 하지 않았지만 바다의 환경정보를 분석하였을 때 추출한 데이터를 기반으로 해파리 출현 확률을 구해주는 기능 한 가지만 추가해주면 되기 때문에 실현가능성은 높다고 생각된다. 2018년 부터 10년단위로 30%씩 피해액을 감축할 수 있다고 예상한다.

#### 3.2) 아이디어의 실현에 따른 파급(기대)효과(효과성)

## 가. 기대효과

2014년, 해파리의 피해를 줄이고자 수억원을 들여 스마트부이를 만들었지만 경남 마산의 시범운영 이후로 좀처럼 확대되지 못하고 있다. 앞의 내용과 같이 해파리 출몰 분석 기술은 기존의 스마트부이에 추가만 해도 사용 가능한 저비용 고효율의 방안이다. 따라서 많은 예산을 사용하지 않고도 시행을 할 수 있는 장소를 점차 늘릴 수 있다. 또한 국제적으로 냉각수를 해수로 사용하는 발전소들이 해파리로 인해 많은 경제적인 피해를 입고 있어, 수출로 인한 경제적인 효과를 가져다 줄 것으로 예상된다.

## 나. 활용계획



스마트부이로 해파리 출현을 모니터링 하면서 바다의 환경을 계속 데이터 기반으로 수온, 염분, pH 등 16개의 조건을 체크하고 있다. 체크한 데이터를 비교하여 해당 해양마다 각 해파리의 출현확률을 기지국에 보낸다. 보낸 데이터를 웹으로 받아 실시간으로 해파리의 출현 정보를 나타내어 사람들에게 제공한다. 추가적으로 해파리 출현 확률이 일정 기준 이상일 경우 알림 서비스를 제공하여 어업종사자 및 발전소 관계자들의 어업활동과 해수를 이용한 냉각수 사용에 피해를 줄일 수 있고, 해수욕장을 찾은 관광객의 피해를 줄이는 등 다양한 방면으로 활용할 수 있다.

### 2.3) (자유타이틀 기재)

