

표지(1)

1. 개요 (1)

2. 분석 목적 (2)

3. 데이터 수집/처리 (2)

4. 데이터 분석 (3-4)

5. 분석 결과(시각화) (1)

6. 기대 효과 / 활용 계획(2)

## 2. 분석목적



어업

매번 2톤 가량의 해파리가 그물에 들어와 어업 피해가 이만저만이 아닙니다.



발전

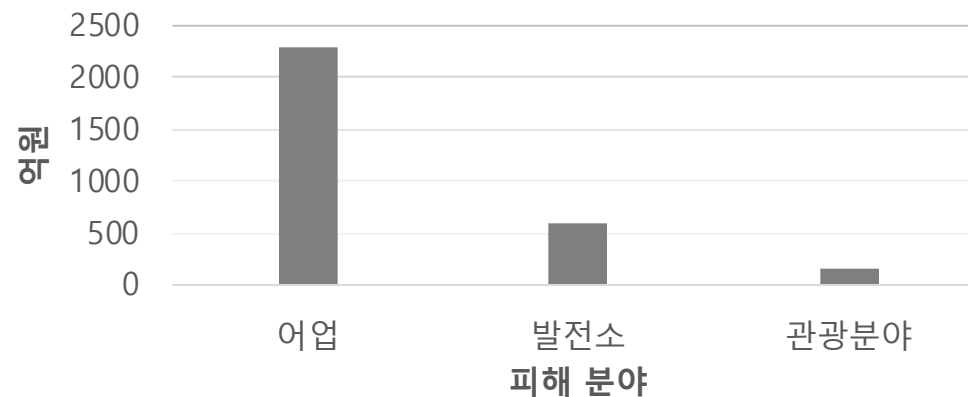
해파리가 원자력 발전소의 취수구를 막는 사례가 계속 늘어나고 있습니다.



관광분야

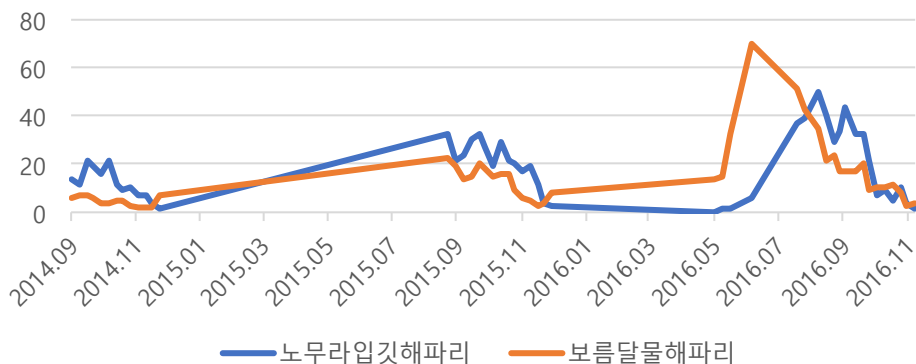
작년에 바다에 놀러갔다가 저희 아들이 해파리에 쏘였어요!

해파리로 인한 2016 피해액



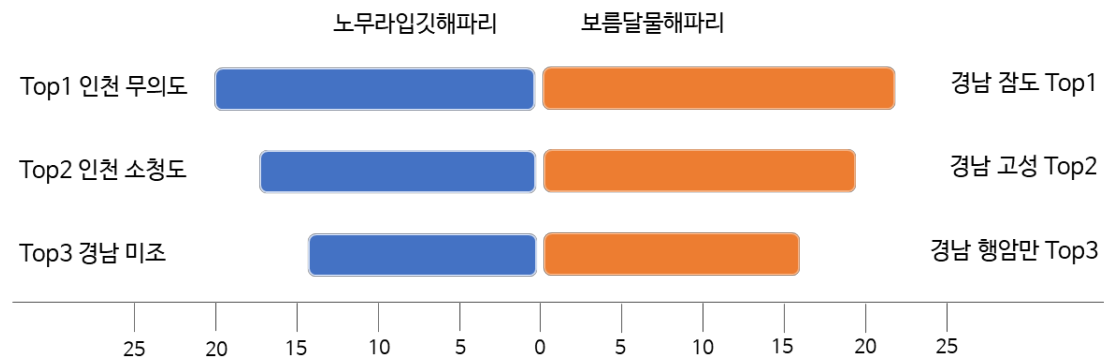
→ 연간 해파리 총 피해액 약 3000억원

기간에 따른 해파리 출현 수



→ 수온 상승으로 인하여 해파리 출현 수 해마다 급증

국내 해파리 피해지역 TOP 3



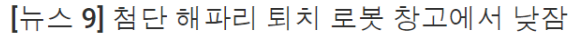
→ 인천과 경남을 비롯한 전 해안지역에서 해파리 출현

## 2. 분석목적



## 스마트부이란?

- ✓ 해상에서 해파리 정보 및 기상정보 등을 관측하는 부표형태의 기구
- ✓ 카메라 등을 통해 해파리 군집을 탐지하여 해파리의 이동 예측 및 방제경로 설정에 활용



등록 2015.06.19 09:16 / 수정 2015.06.19 09:46



## 현재 현황은?

- ✓ 2014년 수익원을 들여 제작

BUT

- ✓ 경남 마산의 시범운영 이후로  
좀처럼 확대되지 못하고 있다

## 기존의 스마트부이

→ 카메라로 실시간 모니터링을 통해  
해파리를 감지하는 시스템

스마트부이에 빅데이터분석을 결합한다면

해파리 출현 지역을 미리 예측할 수 있지

안을까?!



# 3. 데이터 수집/처리

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

## - 해파리 출몰 정보 (Open API)

해파리정보

ENGLISH

(해양수산연구정보) 해파리 출현에 대한 주간 동향, 조치사항, 금후전망, 출현률, 수온, 해파리 종류별 분포도, 주간변동 정보 등 제공

활용신청 (바로그기) 건수 : 66

### Raw Data

jellyFishName	newsPerEnd	newsPerSta	newsSeq	occurrenceSpot
노무라입깃해파리	20161117	20161111	2016112200	○서해 출현     ;저밀도 출현 해역   - 충남 태안 격렬비열도 연안
보름달물해파리	20161117	20161111	2016112200	○남해 출현   일부해역 밀집 출현   ;경남 일부해역 밀집 출현   - 경남 창원 잠도 연안     ;저밀도 출현 해역   - 충남 격렬비열도 연안   - 경남 사량도, 고성 연안   - 부산 광안리 연안
노무라입깃해파리	20161110	20161104	2016111000	○동/서/남해 출현     ;저밀도 출현 해역   - 인천 소야도/백령도 연안   - 전남 소리도 연안   - 경북 영일만 연안
보름달물해파리	20161110	20161104	2016111000	○남해 출현     ;저밀도 출현 해역   - 전남 고흥 득량만 연안   - 경남 잠도 연안   - 부산 놀차만 연안
노무라입깃해파리	20161103	20161028	2016110700	○동/서/남해 출현     ;저밀도 출현 해역   - 충남 녹도 연안   - 전북 심이동파도/위도 연안   - 전남 대흑산도/소리도 연안   - 부산 태종대 연안   - 경북 영일만 연안   - 강원 강릉송정해변/묵호 연안   - 제주 우도 연안

### Data Summary

열이름	정의	추가설명
jellyFishName	해파리이름	총 7종 (노무라입깃해파리, 보름달물해파리 등)
newsPerEnd	조사마감일	
newsPerSta	조사시작일	2014년 10월 ~ 2016 6월
newsSeq	조사주	
occurrenceSpot	출몰지	동,서,남해안 총 155개 지역

## - 어장 환경 모니터링

MEIS  
국가해양환경정보통합시스템

관측정보 | 지리정보 | 지식정보 | 게시판 | 사이트소개 | 정도관리

홈 | 사이트맵 (Home) | Sitemap

출 > 관측정보 > 관측시스템 > 국립수산물과학원 > 어장환경모니터링

관측시스템

어장환경모니터링

자료소개

해역 전체 연안 전체 정점 전체

관측기간 2014. 1. 1. 2016. 12. 31.

텍스트 엑셀 전체

### Raw Data

해역명	연안명	정점명	관측년 도	관측 월	관측 일	Longitud e [degre es_east]	Latitude [degrees _north]	수온(℃)		염분(psu)		pH		용존산소량(mg/ L)		화학산소요구 량(mg/L)	
								표층	저층	표층	저층	표층	저층	표층	저층	표층	저층
서해중부 (VII)	가로림연 안	가로림 01	2016	6	11	126.3111 111	36.99305 556	14.17	14.15	32.03	32.03	7.99	7.97	9.78	9.9	1.53	1.29

암모니아성질 소(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)
표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	
0.009	0.01	0.114	0.142	0.006	0.008	0.129	0.159	0.302	0.28	0.013	0.014	0.028	0.031	0.079	0.092	28.2	20.2	3.23	1.55	1.5

### Data Summary

열이름	추가설명
해역명	총 8개 해역 (서해 중/남부, 남해 동/중/서부, 동해 중/남부, 제주도)
연안명	총 56개 연안 (가로림 연안, 가막만, 김포연안 등)
정점명	총 406개의 정점 (가로림, 진해만, 낙동강 하구 등)
관측날짜	관측년도, 관측월, 관측일 (2014년 10월 ~ 2016년 6월, 21개월)
위도/경도	각 지역의 위도와 경도
해양성분	총 16가지 (수온, 염분, pH, 용존산소량, 화학산소요구량 등)

# 3. 데이터 수집/처리

처리 방법: anaconda jupyter notebook에서 python으로 처리

## - 해파리 출몰 정보 (Open API)

jellyFishName	newsPerEnd	newsPerSta	newsSeq	occurrenceSpot
노무라입깃해파리	20161117	20161111	2016112200	○서해 출현     ;저밀도 출현 해역   - 충남 태안 격렬비열도 연안
보름달물해파리	20161117	20161111	2016112200	○남해 출현   일부해역 밀집 출현   ;경남 일부해역 밀집 출현   - 경남 창원 잠도 연안     ;저밀도 출현 해역   - 충남 격렬비열도 연안   - 경남 사량도, 고성 연안   - 부산 광안리 연안
노무라입깃해파리	20161110	20161104	2016111000	○동/서/남해 출현     ;저밀도 출현 해역   - 인천 소야도/백령도 연안   - 전남 소리도 연안   - 경북 영일만 연안

출몰지(OccurenceSpot)를 각각 한 지역으로 나누어 구글맵(Google Map)을 활용하여 각각의 지역의 위도와 경도를 저장하여 데이터 셋 구축.



jellyFishName	newsPerSta	newsPerEnd	density	spot	longitude	latitude
노무라입깃해파리	20141003	20141010	소량	인천 소청도	124.7141	37.77241
노무라입깃해파리	20141003	20141010	소량	인천 연평도	125.6549	37.6595
노무라입깃해파리	20141003	20141010	소량	충남 격렬비열도	125.5262	36.62027

- 추가/수정된 열  
출몰지(spot): 기존 출몰지를 한 셀로 나누어 표시  
위도/경도(longitude/latitude): 출몰지의 위도와 경도

## - 어장 환경 모니터링

해역명	연간편	정점명	관측년도	관측월	관측일	Longitude [degrees_east]	Latitude [degrees_north]	수온(℃)		염분(psu)		pH		용존산소량(mg/L)		화학산소 요구량(mg/L)				
								표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	
서해중부(Ⅶ)	가로림연안	가로림 01	2016	6	11	126.311111	36.9930556	14.17	14.15	32.03	32.03	7.99	7.97	9.78	9.9	1.53	1.29			
암모니아성 질소(μg/L)		아질산성 질소(μg/L)		질산성 질소(μg/L)		용존무기 질소(μg/L)		총질소(μg/L)		용존무기 인(μg/L)		총인(μg/L)		규산규소(μg/L)		부유물질(mg/L)		클로로필 a(μg/L)		투명도(m)
표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	표층	저층	
0.009	0.01	0.114	0.142	0.006	0.008	0.129	0.159	0.302	0.28	0.013	0.014	0.028	0.031	0.079	0.092	28.2	20.2	3.23	1.55	1.5



...

용존무기인(μg/L)		총인(μg/L)		규산규소(μg/L)		부유물질(mg/L)		클로로필 a(μg/L)		투명도(m)	노무라입깃해파리	보름달물해파리
표층	저층	표층	저층	표층	저층	표층	저층	표층	저층			
0.013	0.014	0.028	0.031	0.079	0.092	28.2	20.2	3.23	1.55	1.5	0	0
											1	0
											0	1

+

기존 어장 환경 모니터링과 수정된 해파리 출몰 정보를 합쳐 각 기간에 해파리가 출몰했다면 0(출몰하지 않음), 1(출몰함)으로 변환하여 데이터 셋 구축.

- 추가/수정된 열  
노무라입깃해파리: 출몰하지 않았다면 0, 출몰했다면 1.  
보름달물해파리: 출몰하지 않았다면 0, 출몰했다면 1.

기타) 해파리는 주로 표층에 분포하므로 저층은 활용하지 않음

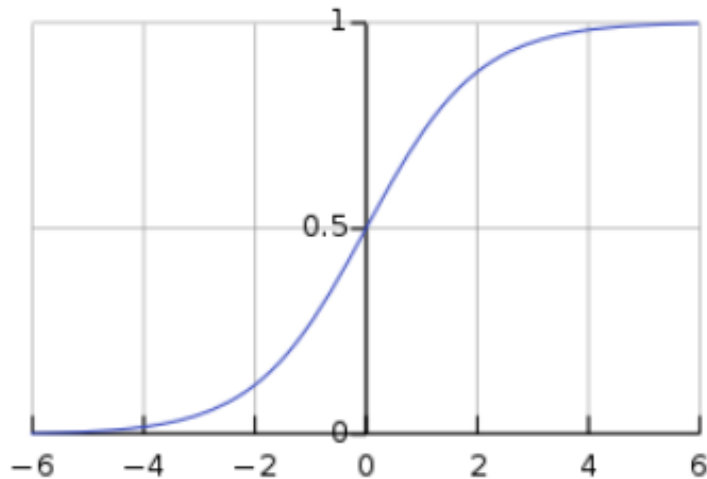
# 4. 데이터 분석

분석 방법: pytho의 sklearn, statmodels을 활용 / 분석 모델 평가를 위해 데이터 셋의 30%를 test data로 추출

분석 모델:

- 로지스틱 회귀분석 (Rogistic Regression)

시그모이드 함수 그래프



함수 수식

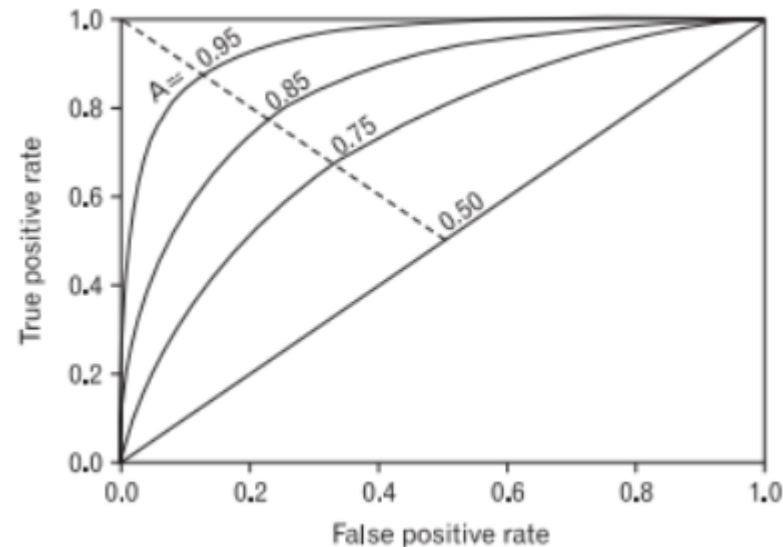
$$y = \frac{1}{1 + e^{-x}}$$

로지스틱 회귀의 목적은 일반적인 회귀 분석의 목표와 동일하게 종속 변수와 독립 변수간의 관계를 구체적인 함수로 나타내어 향후 예측 모델에 사용.

즉, 종속변수에 따라서 독립변수인 0,1를 예측하는 분석모델

- ROC 곡선 (Receiver Operation Characterisitic Curve)

FPR과 TPR을 각각 x,y축으로 놓은 그래프



TPR : True Positive Rate (=민감도, true accept rate)  
1인 케이스에 대해 1로 예측한 비율.

FPR : False Positive Rate (=1-특이도, false accept rate)  
0인 케이스에 대해 1로 잘못 예측한 비율.

# 4. 데이터 분석

## 해양성분과 노무라입깃해파리의 로지스틱 회귀분석 (1차)

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.:	0.09535			
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
OD	-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
BRI/Ca	-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 성분 변수 (출몰에 영향을 주는 변수)

HN3N(암모니아성질소), NH2N(아질산성질소), DIN(용존무기질소), Transparency(투명도)

출몰과 양의 관계인 해양 성분 변수

- HN3N(암모니아성질소),
- NH2N(아질산성질소),
- Transparency(투명도)

출몰과 음의 관계인 해양 성분 변수

- DIN(용존무기질소)

## 해양성분과 보름달물해파리의 로지스틱 회귀분석 (1차)

Logit Regression Results							728
Dep. Variable:		보름달물해파리		No. Observations:			
Model:		Logit		Df Residuals:		711	
Method:		MLE		Df Model:		16	
Date:		Thu, 31 May 2018		Pseudo R-squ.:		0.09535	
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	
OD	-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	
BRI/Ca	-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 성분 변수 (출몰에 영향을 주는 변수)

DO(용존산소량), DIP(용존무기인), SiO2(규산규소)

출몰과 양의 관계인 해양 성분 변수

- DO(용존산소량)
- SiO2(규산규소)

출몰과 음의 관계인 해양 성분 변수

- DIP(용존무기인)



# 4. 데이터 분석

## 해양성분과 노무라입깃해파리의 로지스틱 회귀분석 (2차)

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.:	0.06542			
Time:	11:12:52	Log-Likelihood:	-111.47			
converged:	True	LL-Null:	-121.88			
		LLR p-value:	0.0006433			
=====						
	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
transparency	-0.3229	0.113	-2.861	0.004	-0.544	-0.102
=====						

해양 성분 변수	해양 성분 한 단위 증가당 출몰할 확률
HN3N (암모니아성질소)	$\exp(-28.5020)$ , 약 100감소
NH2N (아질산성질소)	$\exp(-28.5451)$ , 약 100배 감소
Tansparency (투명도)	$\exp(-0.3229)$ , 약 27배 감소
DIN (용존무기질소)	$\exp(25.6581)$ , 약 139050080979 배 증가

노무라입깃해파리 출현 확률을 구하는 함수식

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

## 해양성분과 보름달물해파리의 로지스틱 회귀분석 (2차)

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
=====						

해양 성분 변수	해양 성분 한 단위 증가당 출몰할 확률
DO (용존산소량)	$\exp(0.4437)$ , 약 156 배 증가
DIP (용존무기인)	$\exp(-27.9823)$ , 약 100 배 감소
SiO2 (규산규소)	$\exp(1.1440)$ , 약 314 배 증가

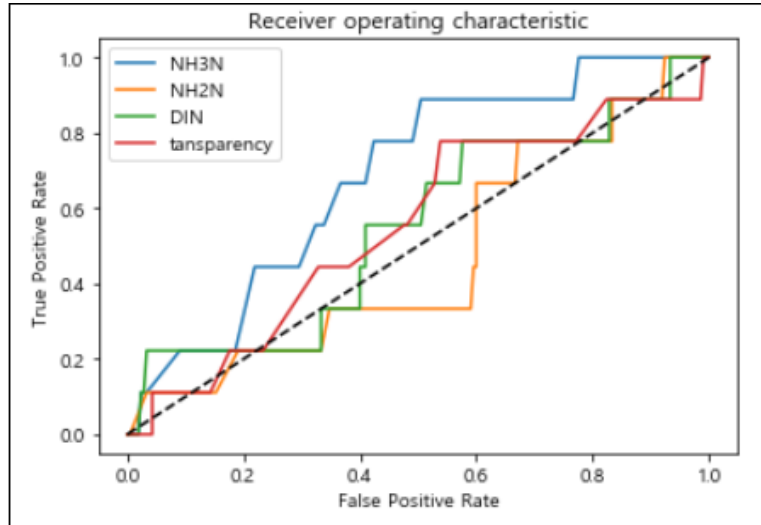
보름달물해파리 출현 확률을 구하는 함수식

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



# 4. 데이터 분석

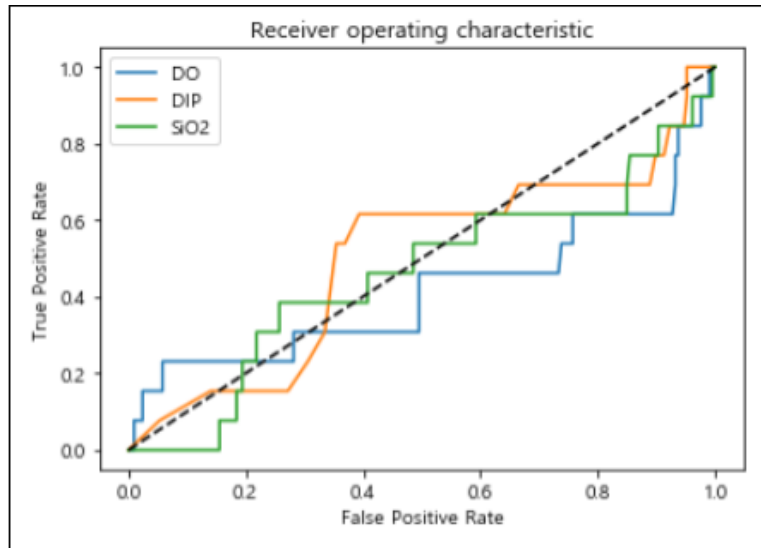
## 해양성분과 노무라입깃해파리의 로지스틱 회귀분석의 ROC 곡선



- 암모니아성질소(NH<sub>3</sub>N)과 투명도(transparency)  
노무라입깃해파리의 출몰 여부를 올바르게 구분해낼 확률이 옳바르지 못하게 구분해낼 확률보다 크다는 점을 확인할 수 있다.
- 아질산성질소(NH<sub>2</sub>N)와 용존무기질소(DIN)  
노무라입깃해파리의 출몰을 옳바르게 예측하지 못한다는 것을 할 수 있다.

즉, 위의 분석 모델에서 NH<sub>3</sub>N과 transparency는 노무라입깃 해파리를 출몰 여부를 파악하는데 유의미한 변수이다.

## 해양성분과 보름달물해파리의 로지스틱 회귀분석의 ROC 곡선

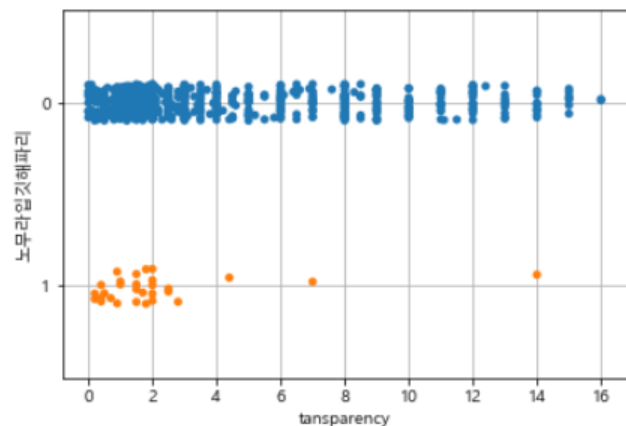


- 용존산소량(DO), 용존무기인(DIP), 규산규소(SiO<sub>2</sub>)  
학습된 모델에서 DO, DIP, SiO<sub>2</sub>가 보름달물해파리 출몰에 영향을 준다는 결과가 나왔지만 옆의 그래프에서는 모델의 성능이 좋지 않다는 결과를 얻을 수 있었다.

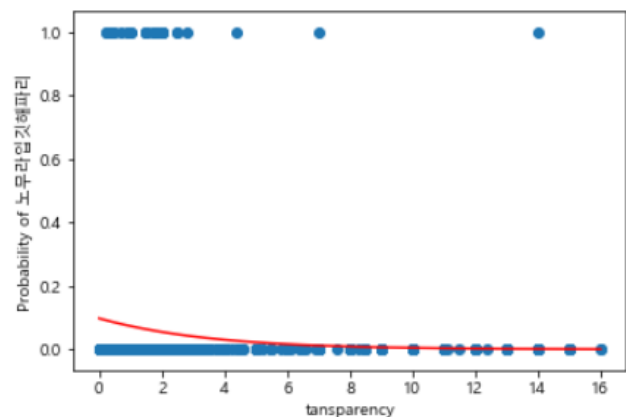
분석 모델의 만들기 위해 더 많은 양의 데이터가 존재한다면 의미있는 결과를 도출할 수 있을 것으로 예상된다.

# 5. 평가 및 한계점

## 노무라입깃해파리와 투명도(transparency)



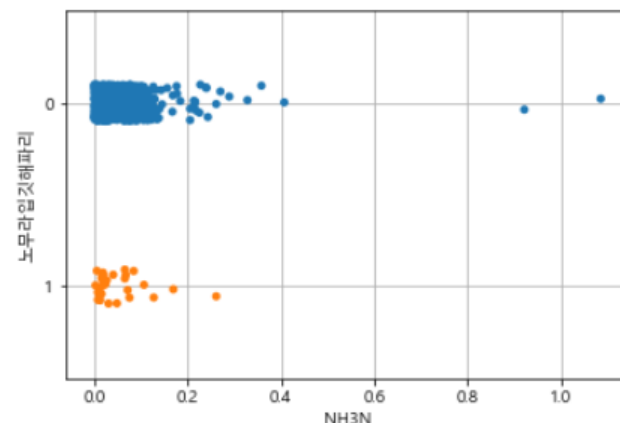
투명도에 따라서  
노무라입깃해파리가  
출몰하는 패턴이  
명확하지는 않지만  
확인 가능함.



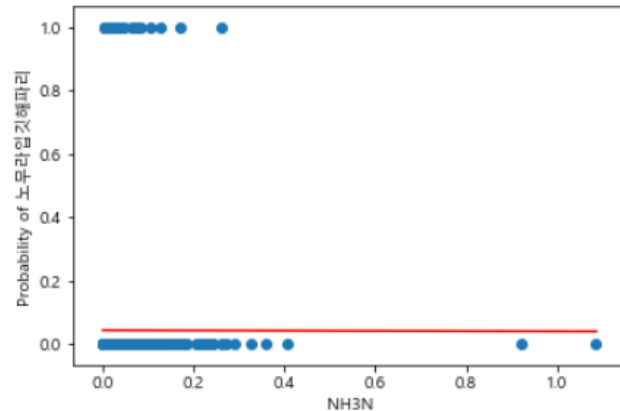
Accuracy 0.960164835165

투명도가 가장 낮을 때,  
노무라입깃해파리를 약  
20% 출몰을 예측해낼 수  
있으며 Test data로 평가  
할 때, 정확도가 약 96%  
이다.

## 노무라입깃해파리와 암모니아성질소(NH3N)



암모니아성질소에  
따라서 노무라입깃  
해파리가 나타나는  
경향성을 확인하기  
어려움



Accuracy 0.960164835165

암모니아성질소로  
노무라입깃해파리를  
구분해낼 수 없음.  
따라서 정확성도  
무의미함

평가) 해파리 연구가 쉽지 않다는 점에서 노무라해파리와 투명도의 관계를 파악함에 따라 앞으로 해파리 연구 이바지할 수 있음  
한계점) 어장 환경모니터링 데이터가 2년치만 존재하여 데이터의 양이 적어져 분석의 정확도를 높이는 데 한계가 있음

# 6. 기대효과/활용계획



관측정보

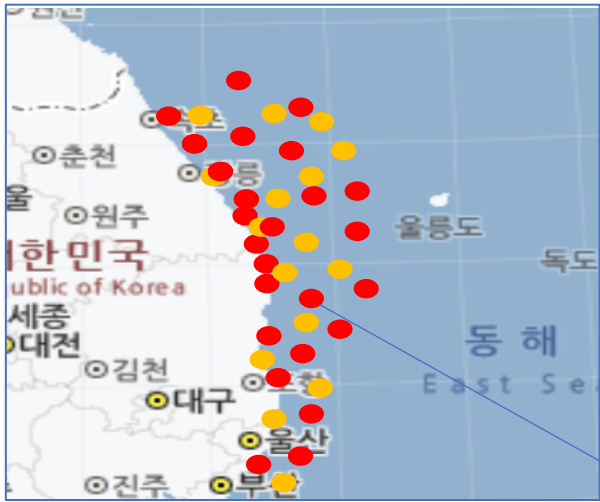
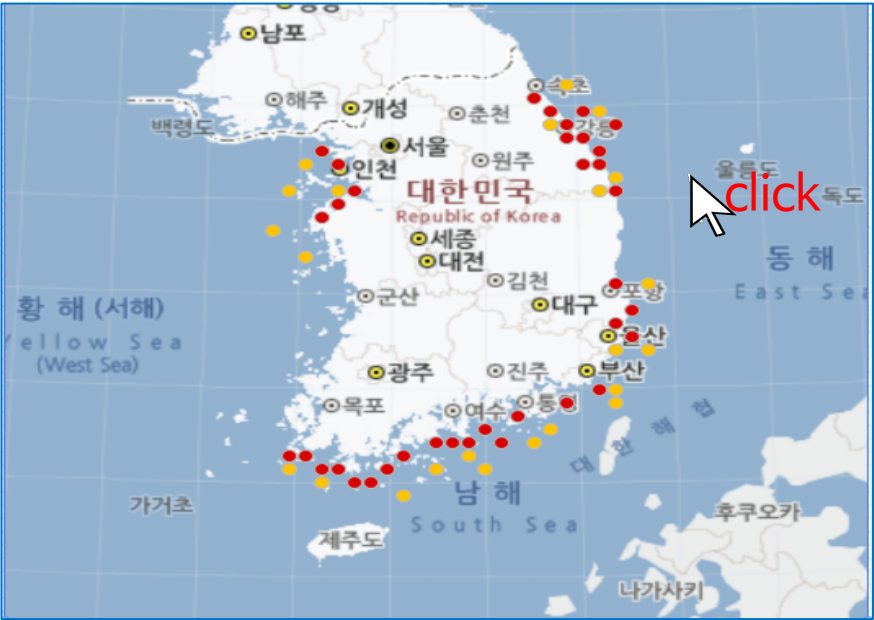
지리정보

게시판

사이트소개

공지사항

동해바다에 다수의 해파리 출몰하였음



해파리명: 보름달물 해파리  
출현확률: 70%  
해파리 특징: 접시처럼 편평한 몸체,  
반투명한 우유빛갈 락  
출현시기: 늦은 봄~여름  
쏘임 증상: 독성은 약하나 사람에 따라  
두드러기 등의 반응 보임

연안명	수온	염분	PH	용존산소량	화학산소요구량	암모니아성 질소
양양연안	17.87	33.95	8.25	8.7	1.06	0
아질산성 질소		질산성질소		용존 무기질소	총질소	용존무기인
0.012		0		0.012	0.079	0.005
총인	규산규소		부유물질		클로로필	투명도
0.008	0.106		1.8		0.23	12