



# 기상 데이터 기반의 해양사고 분석과 예측

---

Team  
S4

01

주제 선정 배경과 문제점

02

데이터 처리와 기초 분석

03

세부 분석 과정

04

솔루션 제안

05

기대 효과



01

## 주제 선정 배경과 문제점



**통영서 어선 전복, 강한 바람에 뒤집어져 ... 60척 동원된 수색** 통영 어선 전복 사망자 4명으로 늘어...실종자 4명 수색 계속

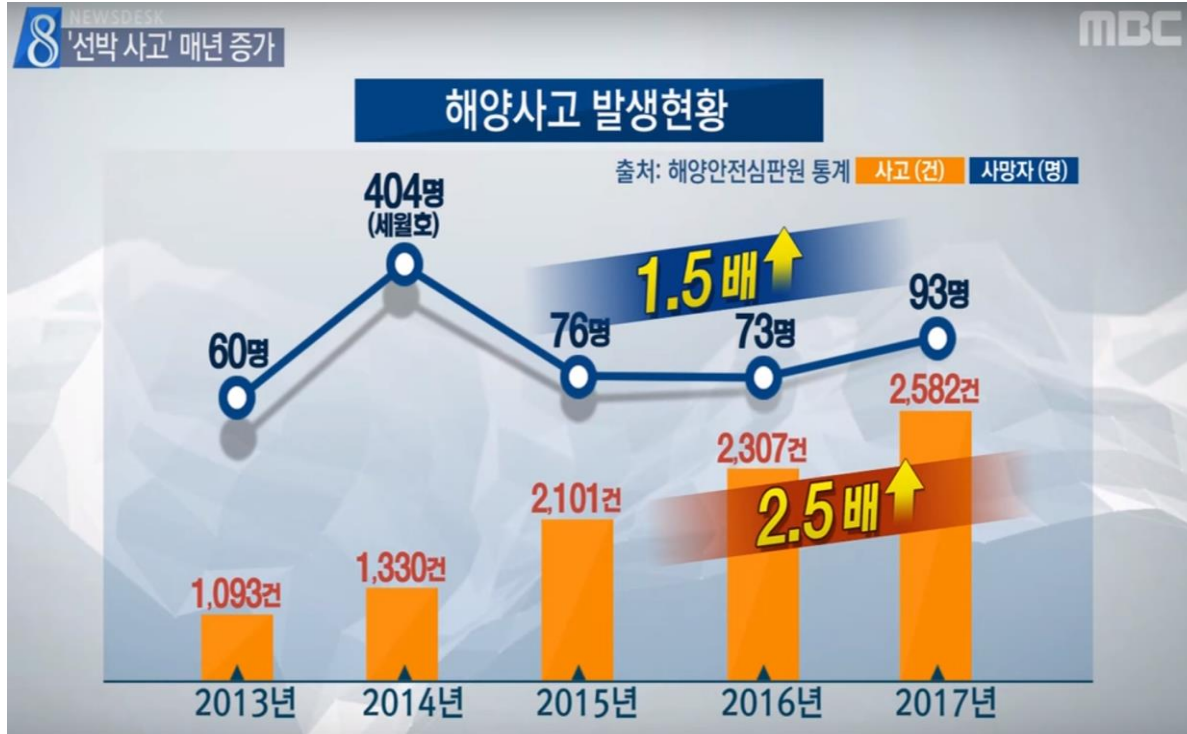
생존자 진술에 따르면, 사고 당시 3m에 이르는 파도로 인해 제일호가 무게중심을 잃고 우 선원 11명이 타고 조업에 나선 제11제일호는 지난 3월 6일 오후 11시께 통영시 욕지면 좌사리도 남서방 해상에서 조업을 마치고 기상악화(풍랑주의보 발효 파고 2~4m, 풍속 13~17m/s)로 인해 안전해역에서 어획물 선별작업을 하기 위해 항해 중 전복됐다.

“ 기상은 지상 사고 뿐만 아니라, **해양사고에도 큰 영향을 미친다** ”



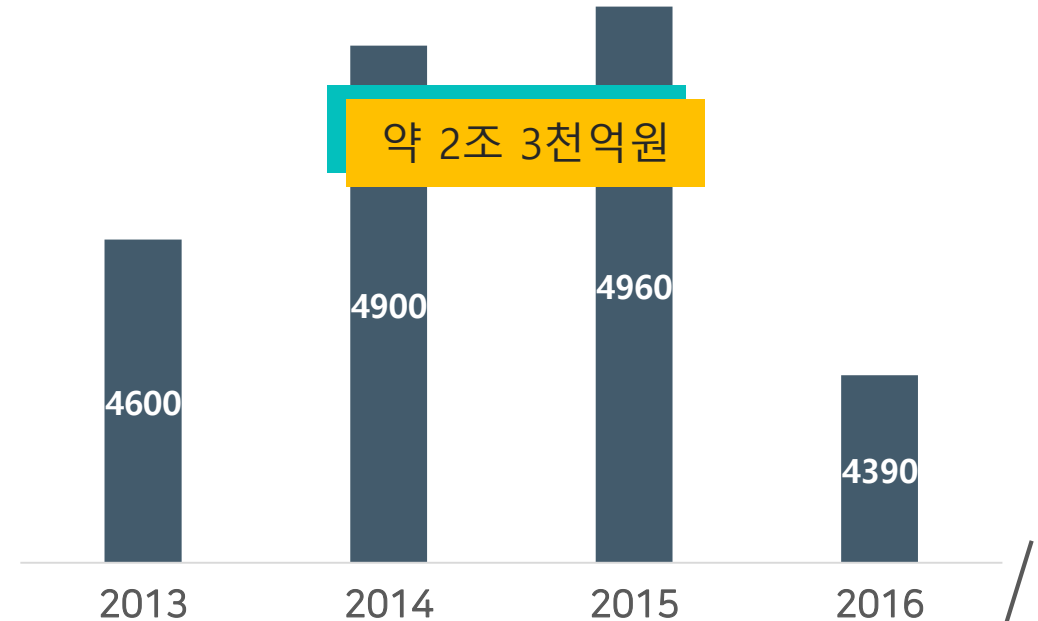
- 01 육지와와의 거리, 사고 대응 시간 증가
- 02 해상에서 사람의 생존능력 저하
- 03 기상 상황에 따른 조치 불가
- 04 다른 교통 수단에 비해 많은 탑승인원
- 05 현장 보존 불가능에 따른 대처 한계

“ **해양사고는 일반적 교통 사고보다 위험성이 높으며, 근본적인 사고의 발생을 방지하는 것이 최우선시 되어야 한다** ”



해양사고로 인한 재산피해액

단위 : 억(원)



더 이상

해양사고와 사망자의 발생을 간과할 수 없으며, 이에 대한 분석이 시급하다

# 어떻게 분석할 것인가?

- 1) 사고 데이터를 통한 탐색적 분석
- 2) 학습을 통한 사고 발생 예측



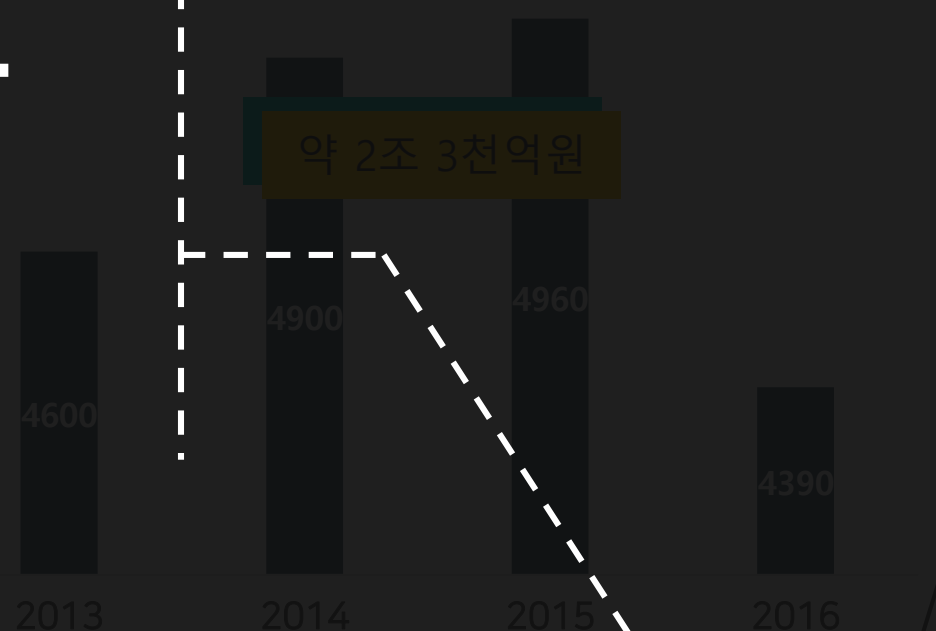
## 솔루션 제안과 기대효과 산출

더 이상

해상사고와 사망자의 발생을 간과할 수 없으며,

이에 대한 분석이 시급하다

해상사고로 인한 재산피해액  
단위 : 억(원)



02

## 데이터 처리와 기초 분석



## 해양사고 상세데이터 | 2011~2017

발생일시	관찰해경서	발생해역	발생유형
발생원인	기상특보	사고선박수	구조인원
사상 및 실종자	사고위치	선종	톤수

(출처 : 공공데이터 포털)

## 해양 기상부이 데이터 | 2011~2017

풍향(deg)	풍속(m/s)	기압(hPa)	최대파고(m)
유의파고(m)	파향(deg)	평균파고(m)	파주기(sec)
습도(%)	수온(°C)	기온(°C)	

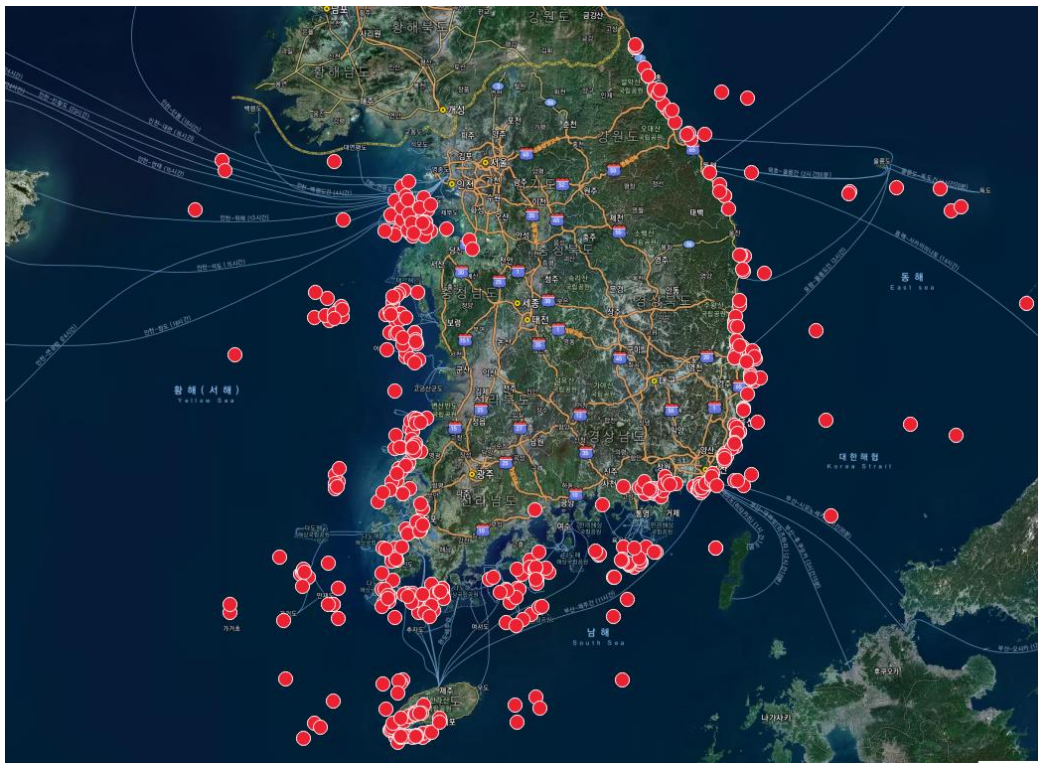
(출처 : 기상자료개방 포털)

## 조위관측소 기상데이터 | 2011~2017

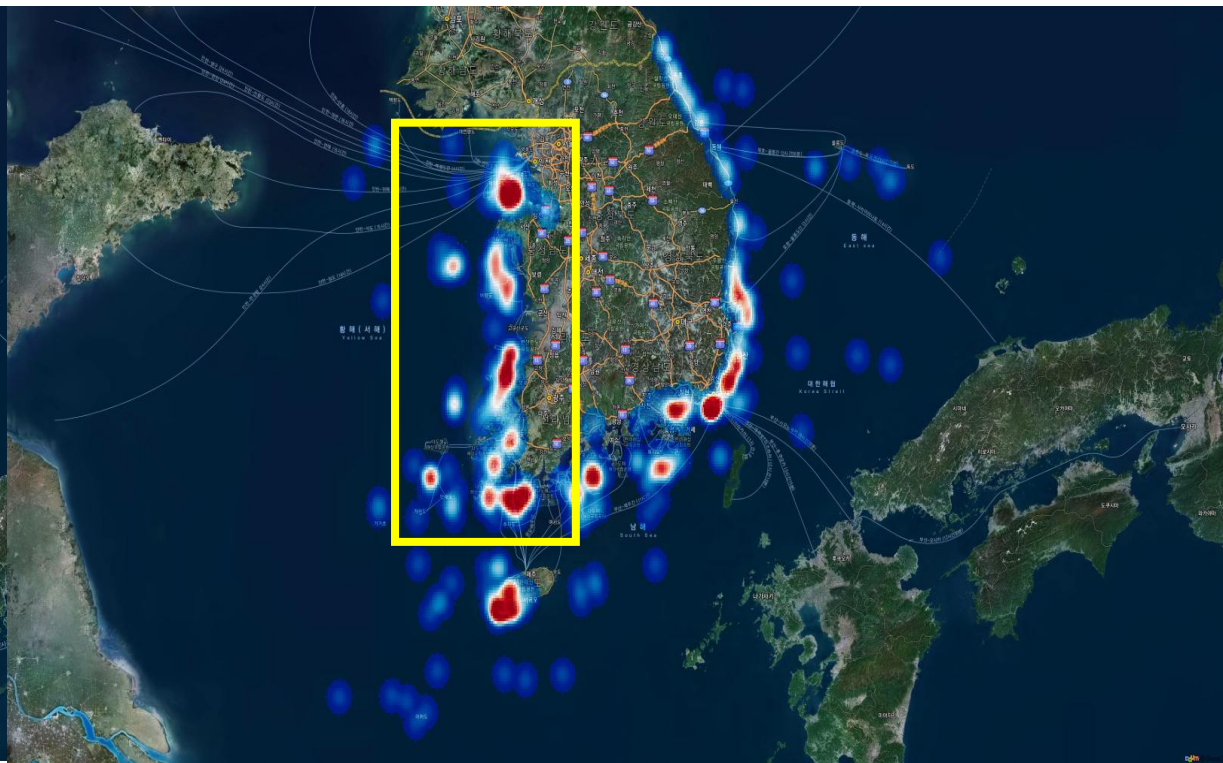
수온	풍속	풍향	기온
기압	조위(cm)		

(출처 : 국립해양조사원)

해양사고 발생지점

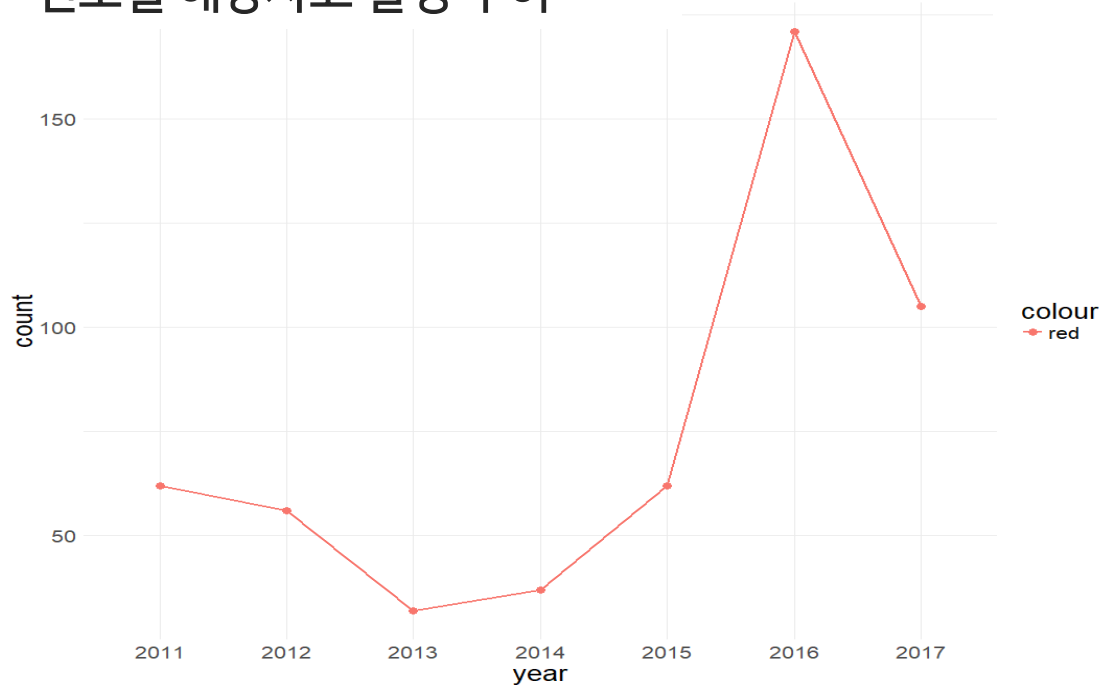


밀도 기반의 히트맵 분포 시각화



전체 기간의 데이터 속에서 발생한 사고 중, 기상악화는 총 574건

연도별 해양사고 발생 추이



월별 해양사고 발생 추이

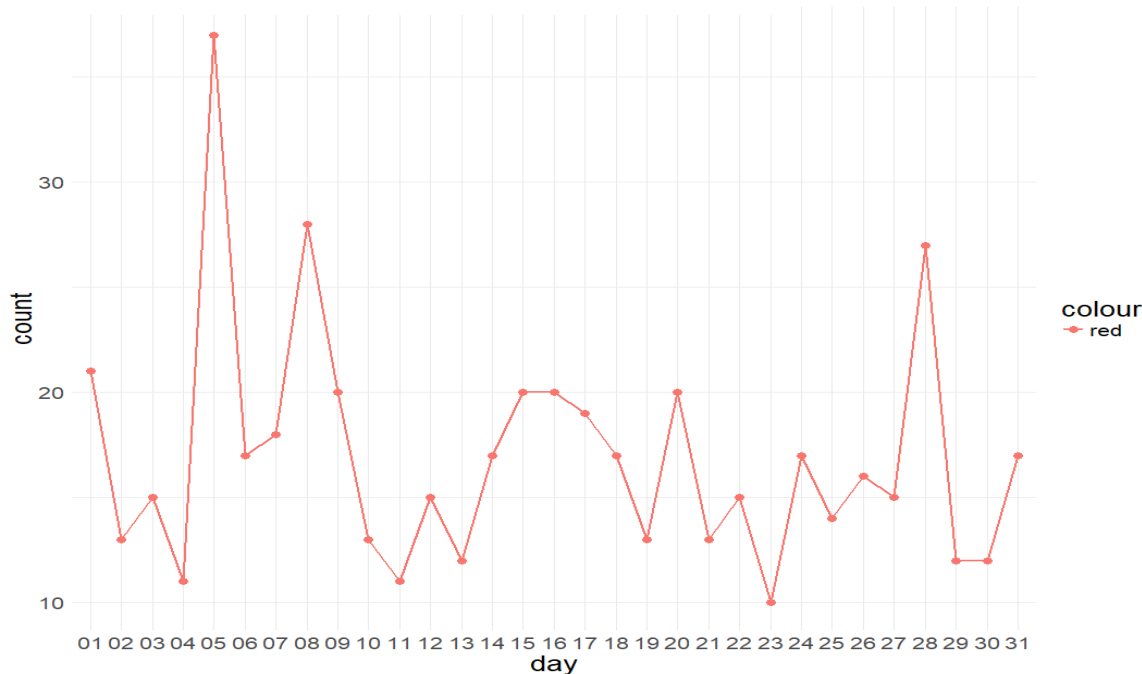


“

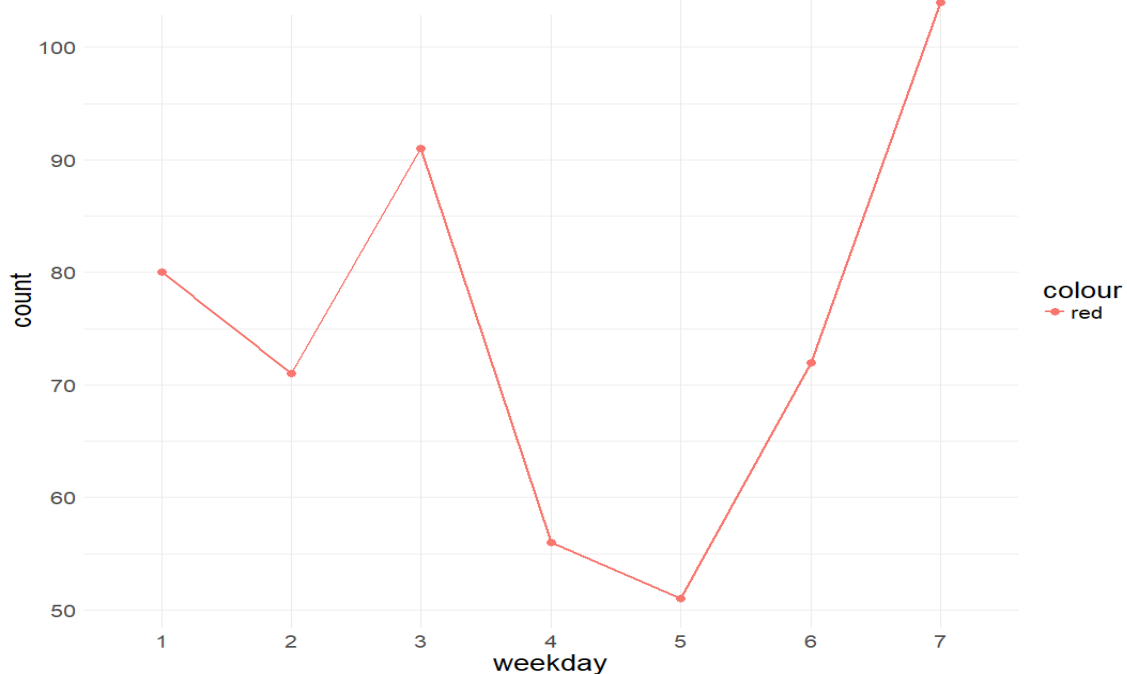
해상사고는 대체로 우상향의 추세에 있으며,  
4월과 8월, 10월에 높은 빈도를 보인다

”

각 월의 일별 해양사고 발생 추이

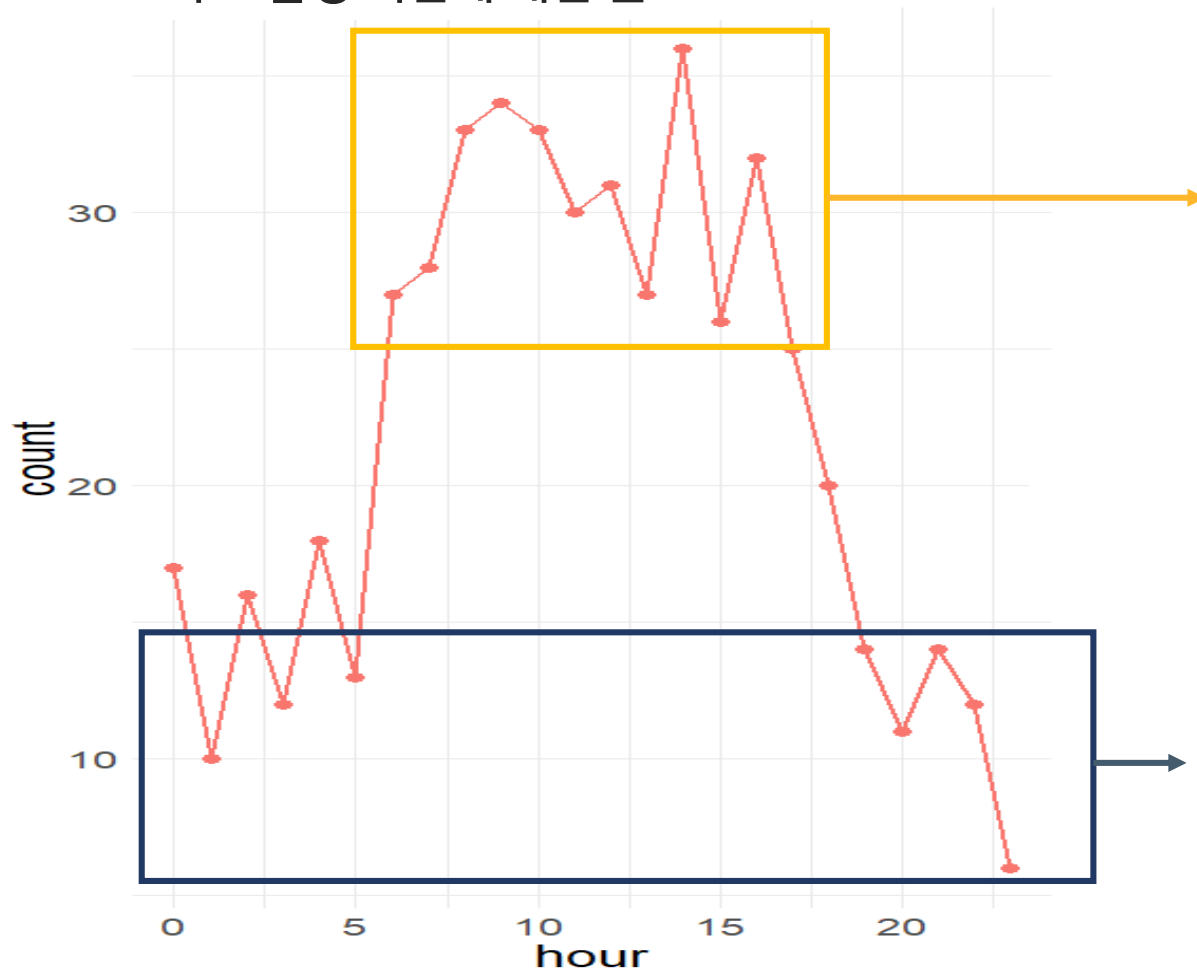


요일별 사고 발생 빈도



“ 해양사고는 비교적 월초와 월말에 더 많이 발생했고, 수요일과 일요일에 다른 요일보다 많이 발생했다 ”

사고 발생 시간에 대한 분포

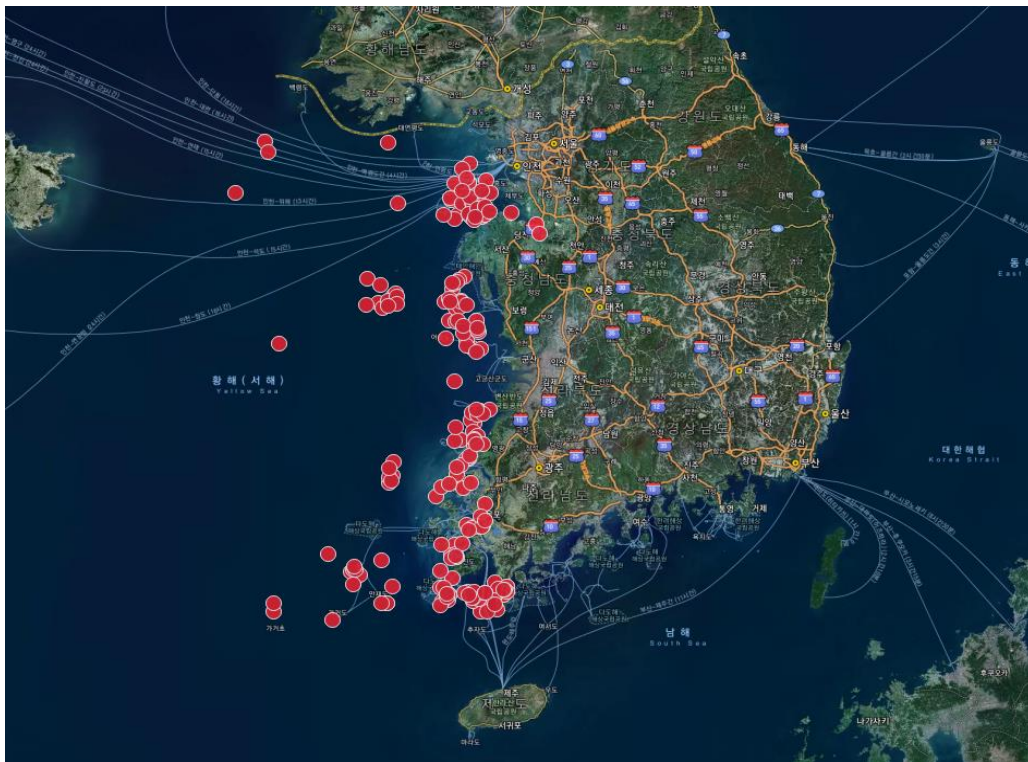


“ 8시부터 16시까지의 사고량이 다른 시간대에 비하여 확연한 발생의 빈도 차이를 보였다 ”

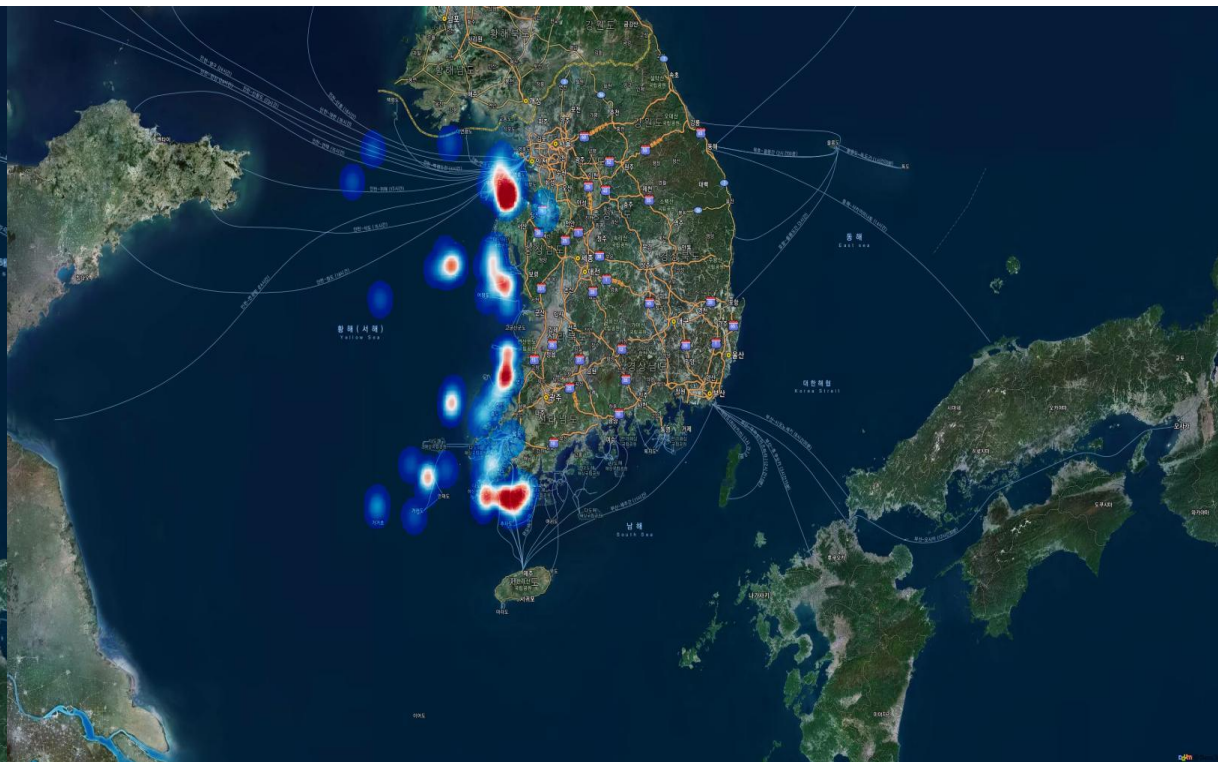
“ 18시부터 익일 5시까지는 사고발생의 빈도가 비교적 낮았다 ”



해상사고 발생지점



밀도 기반의 히트맵 분포 시각화



사고에 대한 지역의 영향을 내포하기 위해 분석 범위 구분  
본 연구에선, 사고가 밀집되어 있는 **서해 부근**을 분석 범위로 선정



### 해양기상 부이 지도 맵핑

해양사고와 기상과의 관계를 파악하기 위하여 기상 데이터 활용  
지도에 해양 기상부이를 맵핑



### 조위관측소 지도 맵핑

해양사고 발생지점과 해양 기상부이와의 거리에서 괴리를 줄이기 위해  
조위관측소 데이터를 추가 활용하여 맵핑



### 해양사고 지도 맵핑

분석의 범위로 선정했던 서해에서 발생한 해양사고 데이터를 맵핑



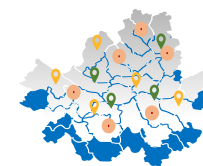
### 거리행렬 최근린 분석

해양사고 지점에 해당하는 기상을 결합하기 위하여 GIS의 거리행렬 최근린 분석 사용  
- 각 사고는 가장 가까운 거리의 기상 관측 지점으로 귀속



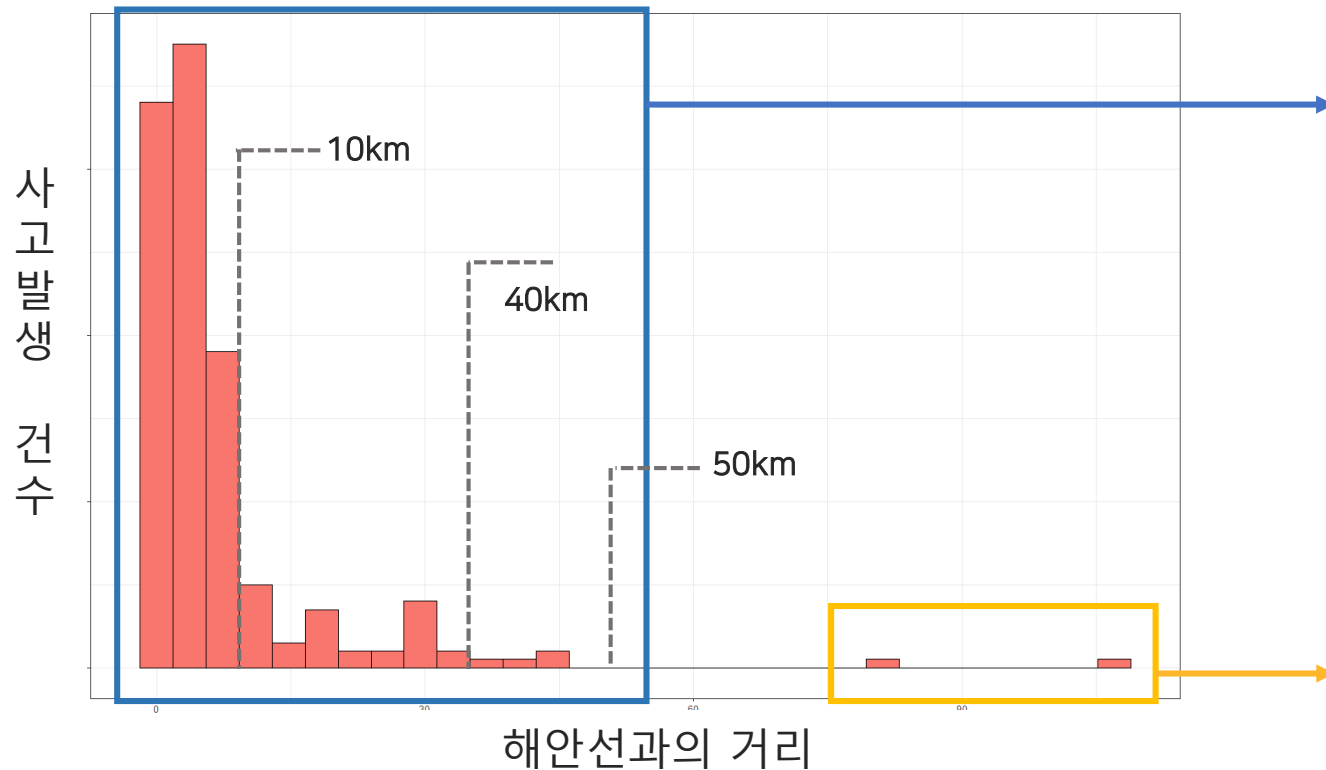
### 사고 데이터와 기상 병합

관측소에 귀속된 해양사고 발생 건은 해당 기상과 병합하게 됨  
- 기상의 측정 단위는 모두 '시간' 단위로, 가장 정밀한 시점의 데이터 사용





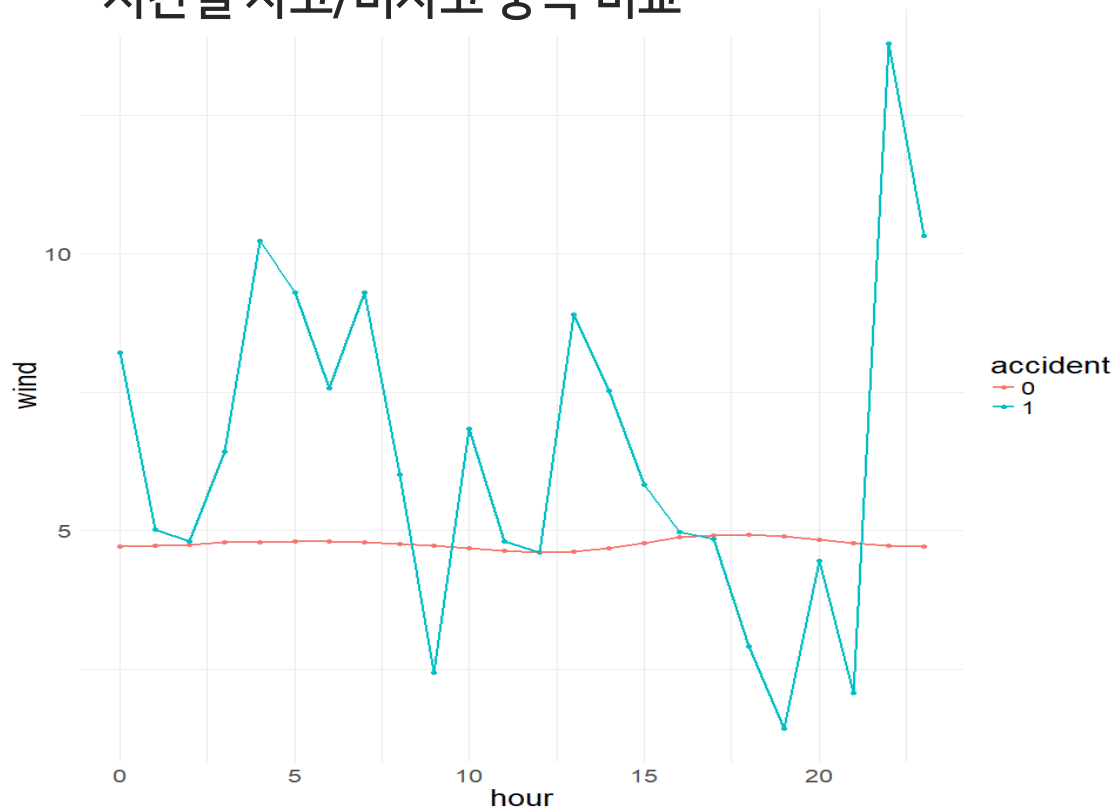
사고 지점과 해안선과의 거리



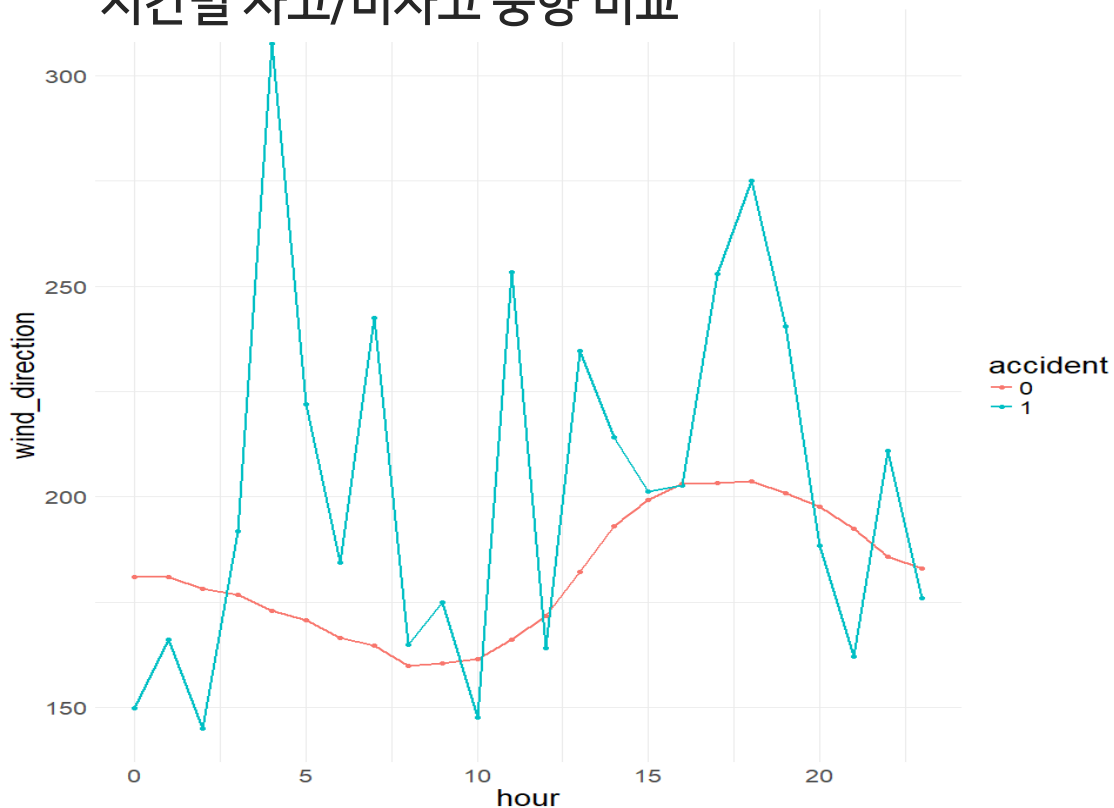
서해 해양사고 데이터 221건 중  
80%가 10km 미만,  
97%가 40km 미만,  
99%가 50km 미만에서 발생

1%에 해당하는 2건의 사고는 해안선과의  
거리가 80km 이상에서 발생  
기상 관측지점과의 거리도 100km 이상으로,  
이상치로 간주하여 본 연구에서 분석 대상 제외

시간별 사고/미사고 풍속 비교

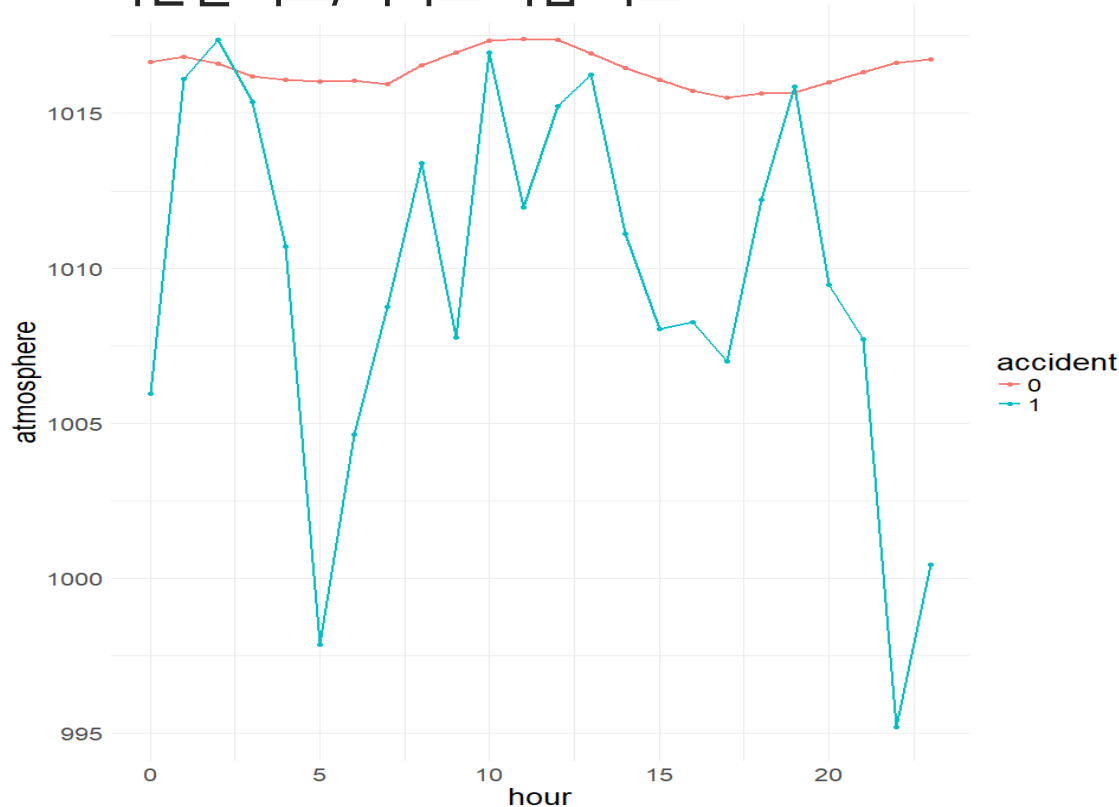


시간별 사고/미사고 풍향 비교

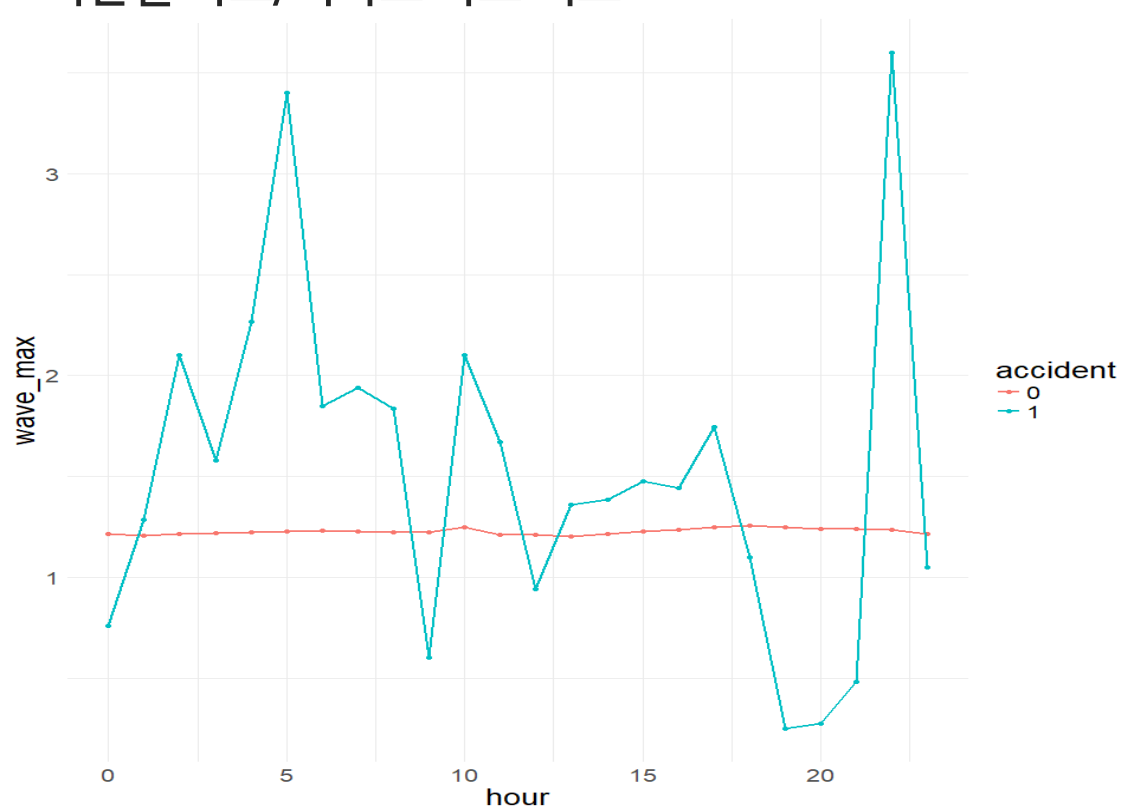


“ 사고 발생 당시는, 사고가 발생하지 않았을 때에 비해 ”  
대체로 높은 풍속을 보였고, 200-300 사이의 풍향이었다

시간별 사고/미사고 기압 비교



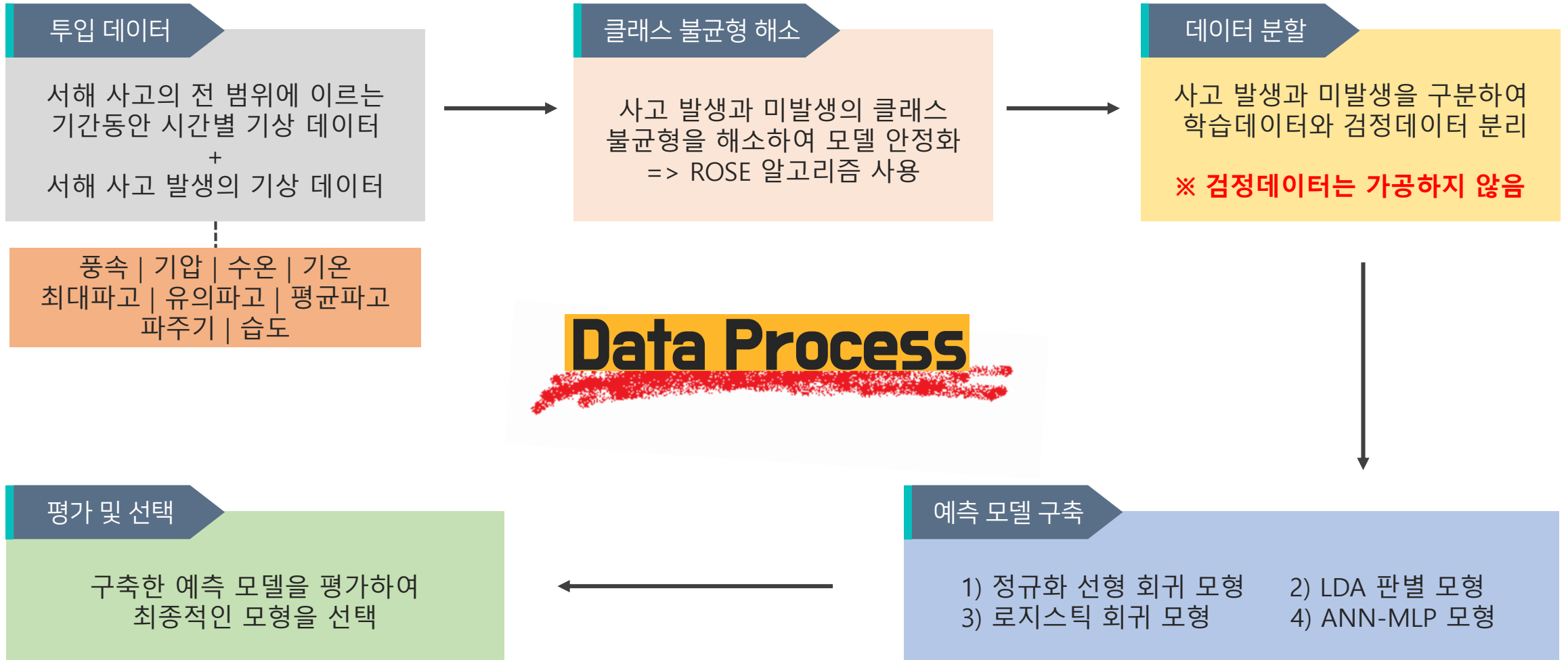
시간별 사고/미사고 파고 비교



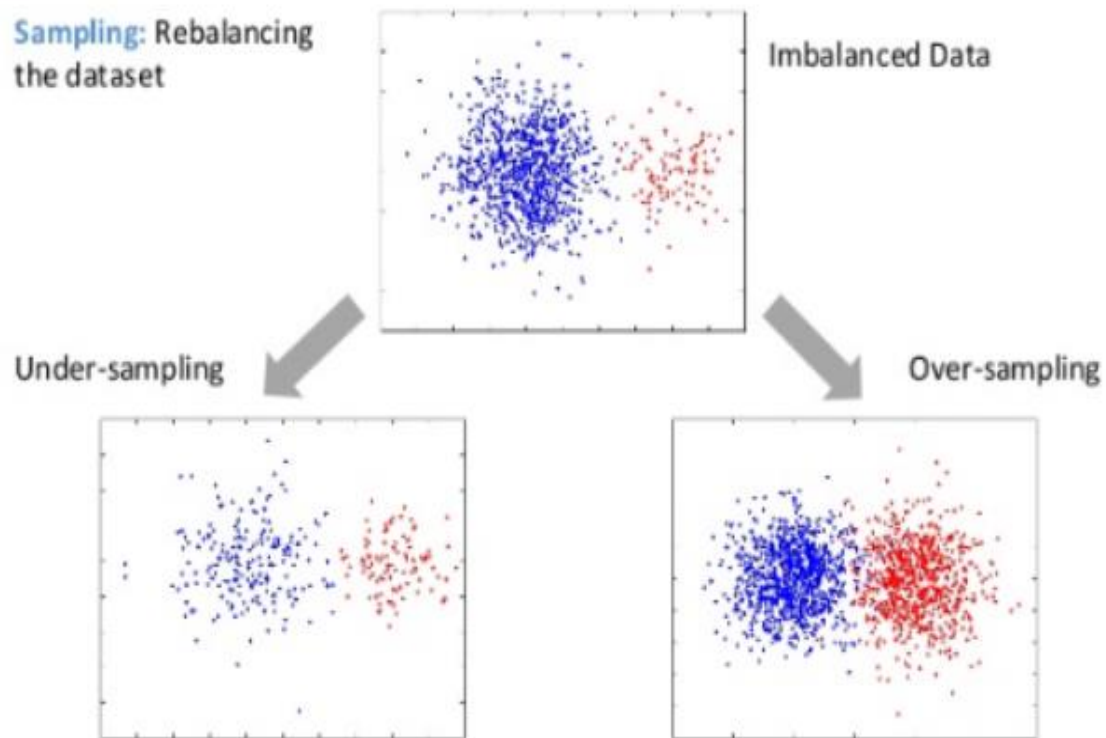
“ 사고 발생 당시는, 사고가 발생하지 않았을 때에 비해 대체로 낮은 기압을 보였고, 파고는 높은 편이었다 ”

03

## 세부 분석 과정



Sampling: Rebalancing  
the dataset



분류 모델에서 클래스 간 과도한 빈도 차이는  
모델의 정확도에 부정적인 영향을 미치며  
데이터 불균형 발생



언더 샘플링은 데이터 손실의 문제를,  
오버 샘플링은 과적합화의 문제를 겪음  
이를 해결하기 위해 ROSE 알고리즘 활용



ROSE는 평활화된 부트스트랩에 따라  
새로운 데이터를 생성하여 학습의 모델  
추정과 정확도 문제를 동시에 해결하는 틀 제공

본 연구에서도, ROSE를 활용하여 학습 데이터 중 사고 데이터의 클래스 균형 조정



### 정규화 선형 회귀 모형

- (1) Lasso 회귀 모형 :  $\alpha = 1$ ,  $\lambda = 0.0009495$
- (2) Ridge 회귀 모형 :  $\alpha = 0$ ,  $\lambda = 0.01198$
- ✓ (3) elastic net 회귀 모형 :  $\alpha = 0.5$ ,  $\lambda = 0.001576$



### LDA 판별 모형

- prior (사전확률) :  $\frac{1}{2}$
- scale (표준화) 적용



### 로지스틱 회귀 모형

- Stepwise Selection : 변수선택법 적용
- AIC를 기준으로 설정



### 인공신경망 다층 퍼셉트론

- hidden layer = 1
- hidden unit = 100
- solver = lbfgs

## Training dataset 8 : Test dataset 2



# Learning . . .

기상 데이터



사고발생 예측

## 모델 평가 지표

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

예측 모델에 대해서 평가할 때 단순히 정확도(accuracy)로 판단하는 것보다는 모델의 특성이 반영된 지표를 사용해야 함

해양사고의 발생을 예측하는 상황에선 실제 사고 중에서, 사고가 발생할 것으로 예측하는 비율이 높은 것이 중요함

심각한 질병을 앓고 있는 사람의 병을 진단하는 것처럼, 예측을 실패했을 경우의 위험 수준이 고려되어야 하기 때문

이에 따라 본 연구는, 예측 모델의 성능을 평가하는 과정에서 실제 발생 중 발생 예측의 비율인 Recall 지표 사용



Model	Accuracy	Recall
Elastic net 회귀	0.6452	0.8965
LDA 판별 모형	0.6454	0.8621
로지스틱 회귀 모형	0.6548	0.7931
인공신경망 다층 퍼셉트론 모형	0.5596	0.6200

최종적인 모델 평가 결과, Elastic net 회귀 모형 채택

## 정규화 선형 회귀

선형회귀 계수(weight)에 대한 제약 조건을  
추가함으로써 모델이 과도하게 최적화되는 현상,  
즉 과최적화를 막는 방법

## 3가지 모델이 존재

- 라쏘(Lasso)

$$\text{Lasso: } \underset{\beta}{\operatorname{argmin}} \sum \varepsilon_i^2 + \lambda \sum |\beta_k|$$

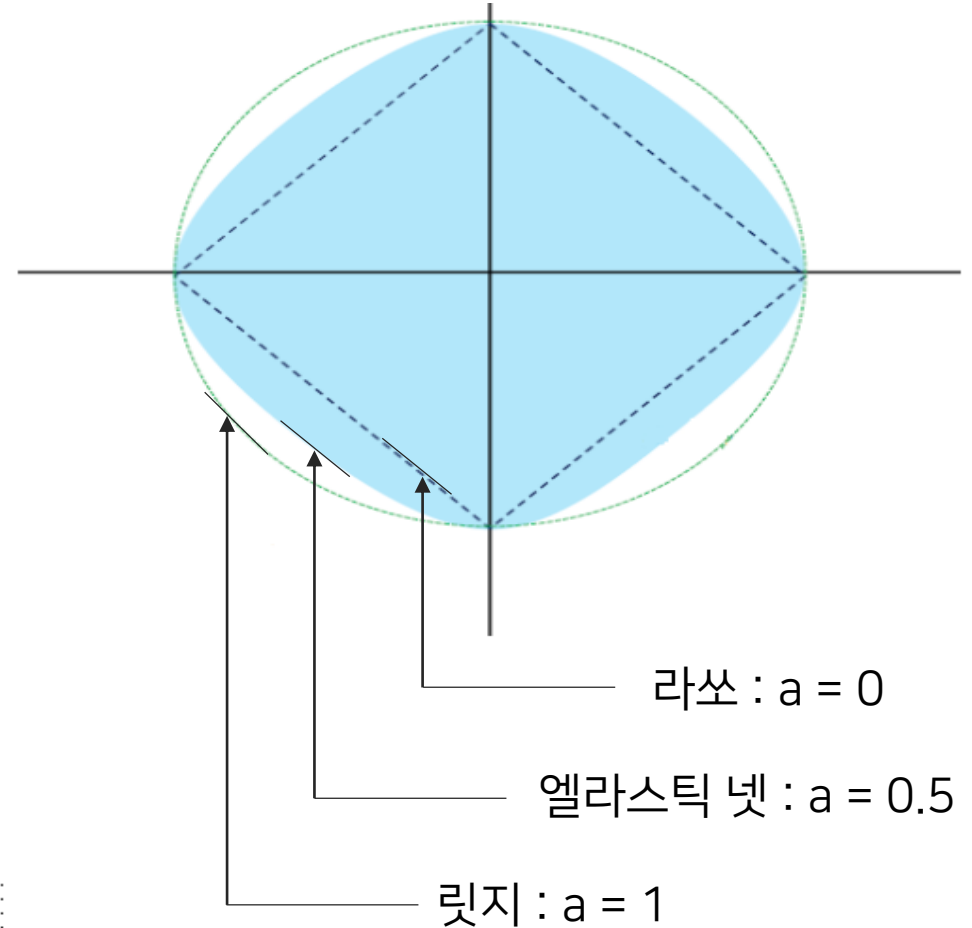
- 릿지(Ridge)

$$\text{Ridge: } \underset{\beta}{\operatorname{argmin}} \sum \varepsilon_i^2 + \lambda \sum \beta_k^2$$

- 엘라스틱 넷(Elastic net)

$$\text{Elastic net: } \underset{\beta}{\operatorname{argmin}} \sum \varepsilon_i^2 + \lambda_1 \sum \beta_k^2 + \lambda_2 \sum |\beta_k|$$

## 수축 패널티 a값에 따른 모형 구분



## 라쏘 회귀

Lasso(Least Absolute Shrinkage and Selection Operator) 회귀 모형은 가중치의 절대값의 합을 최소화하는 것을 추가적인 제약 조건으로 한다.

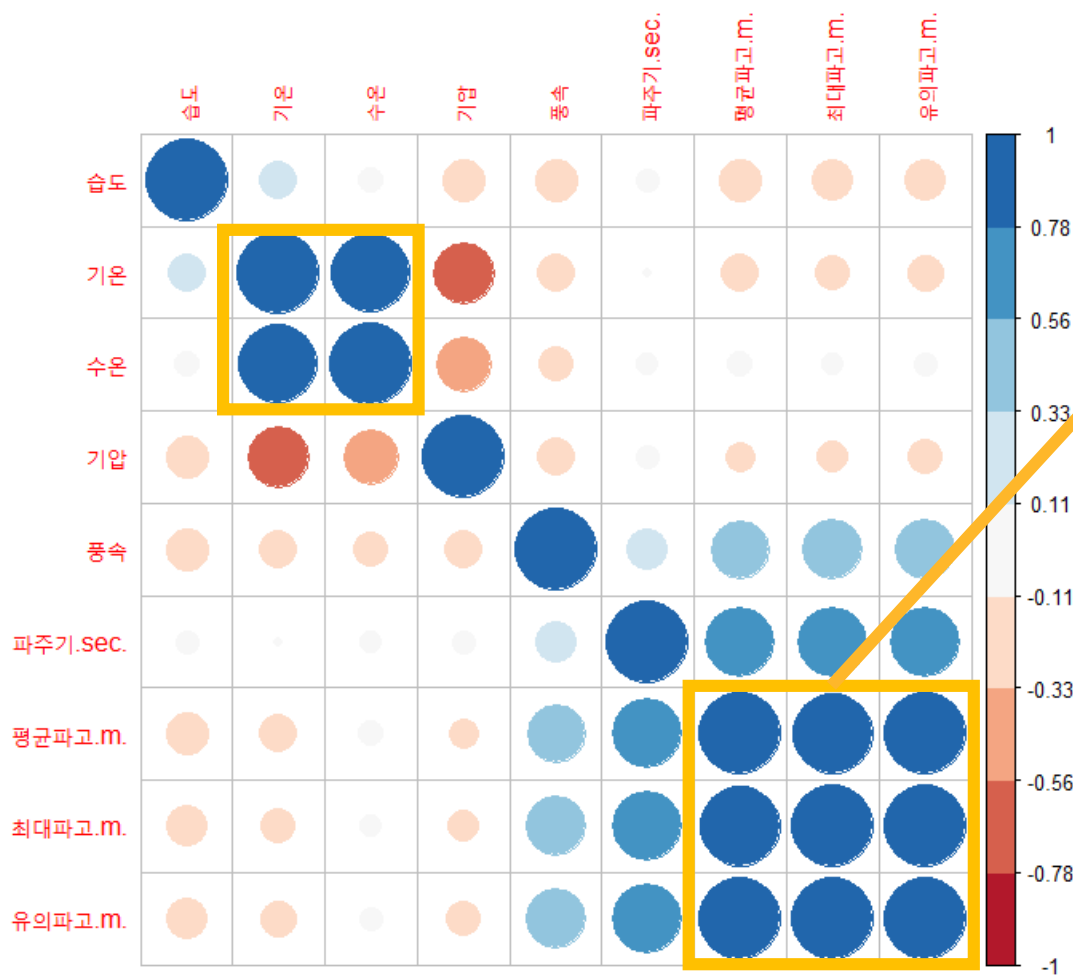
## 릿지 회귀

Ridge 회귀 모형에서는 가중치들의 제곱합(squared sum of weights)을 최소화하는 것을 추가적인 제약 조건으로 한다.

## 엘라스틱 넷 회귀

Elastic Net 회귀 모형은 가중치의 절대값의 합과 제곱합을 동시에 제약 조건으로 가지는 모형이다.  
 $\lambda_1$ ,  $\lambda_2$  두 개의 하이퍼 모수를 가진다. / 하이퍼 모수 : 잔차 제곱합과 제약 조건 비중 조절 모수

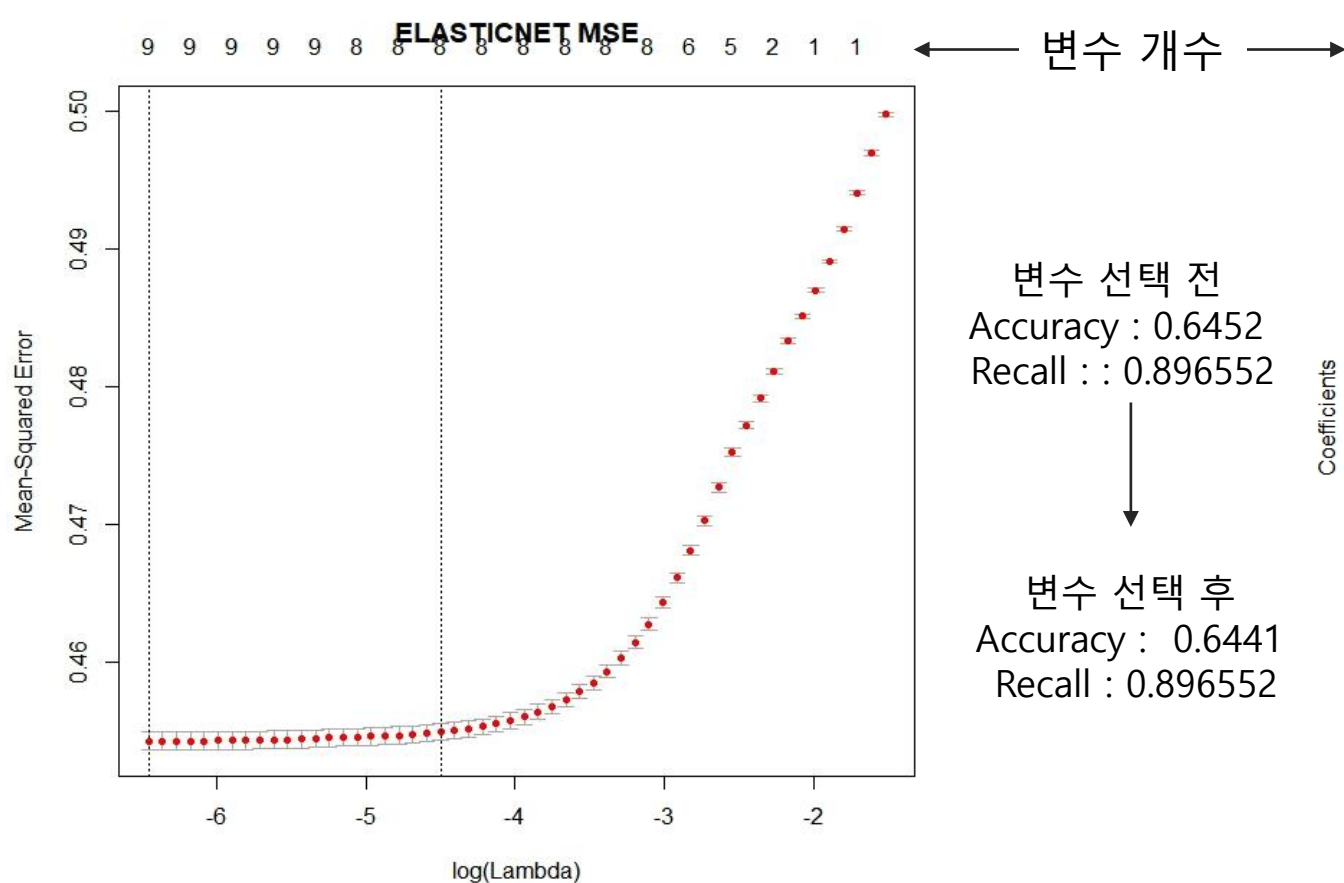
Model	변수선택	특징
라쏘 회귀 모형	불가능	상관관계 높으면, 크기 큰 변수 우선감소
릿지 회귀 모형	가능	상관관계 높으면 성능 저하
엘라스틱 넷 회귀 모형	가능	상관관계 큰 변수 동시 선택/배제 조절 최적화



파고와 관련된 3개의 기상들이 서로 90%에 가까운 상관관계를 가짐

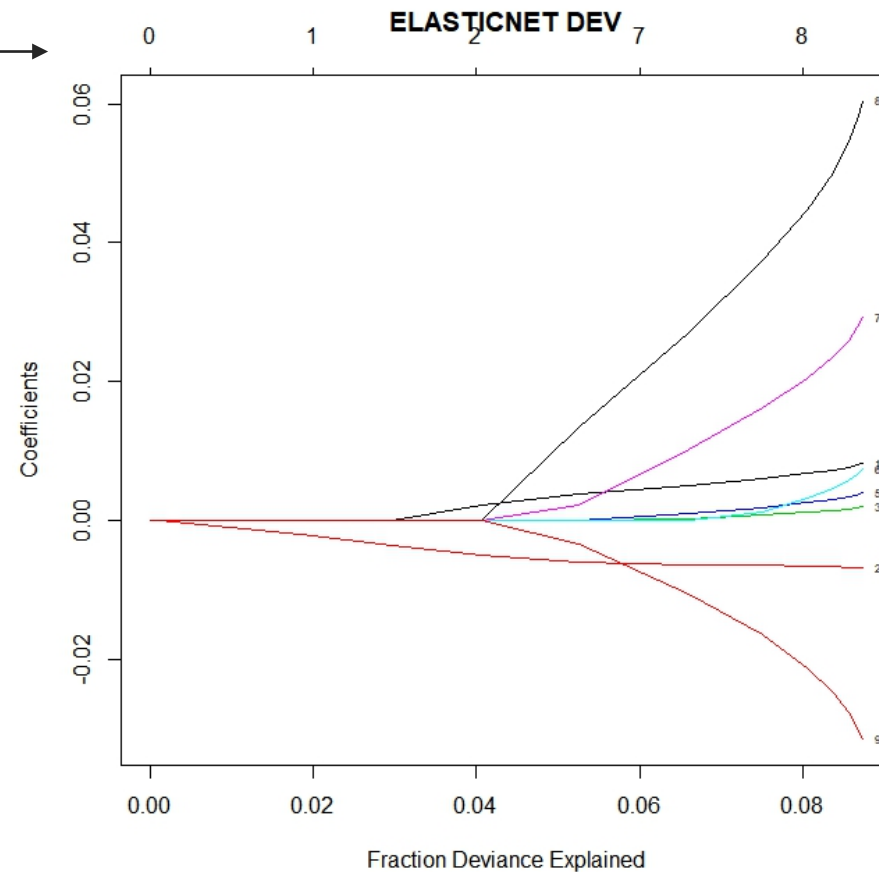
상관관계 속에서 발생하는 분산팽창요인(다중공선성)을 엘라스틱 넷의 장점인 변수 선택/배제 최적화를 활용

도출되는 엘라스틱 넷의 모형에서는 변수들이 가지는 상관관계의 영향이 최적화(정규화)되어 가장 높은 성능의 모델이 생성됨



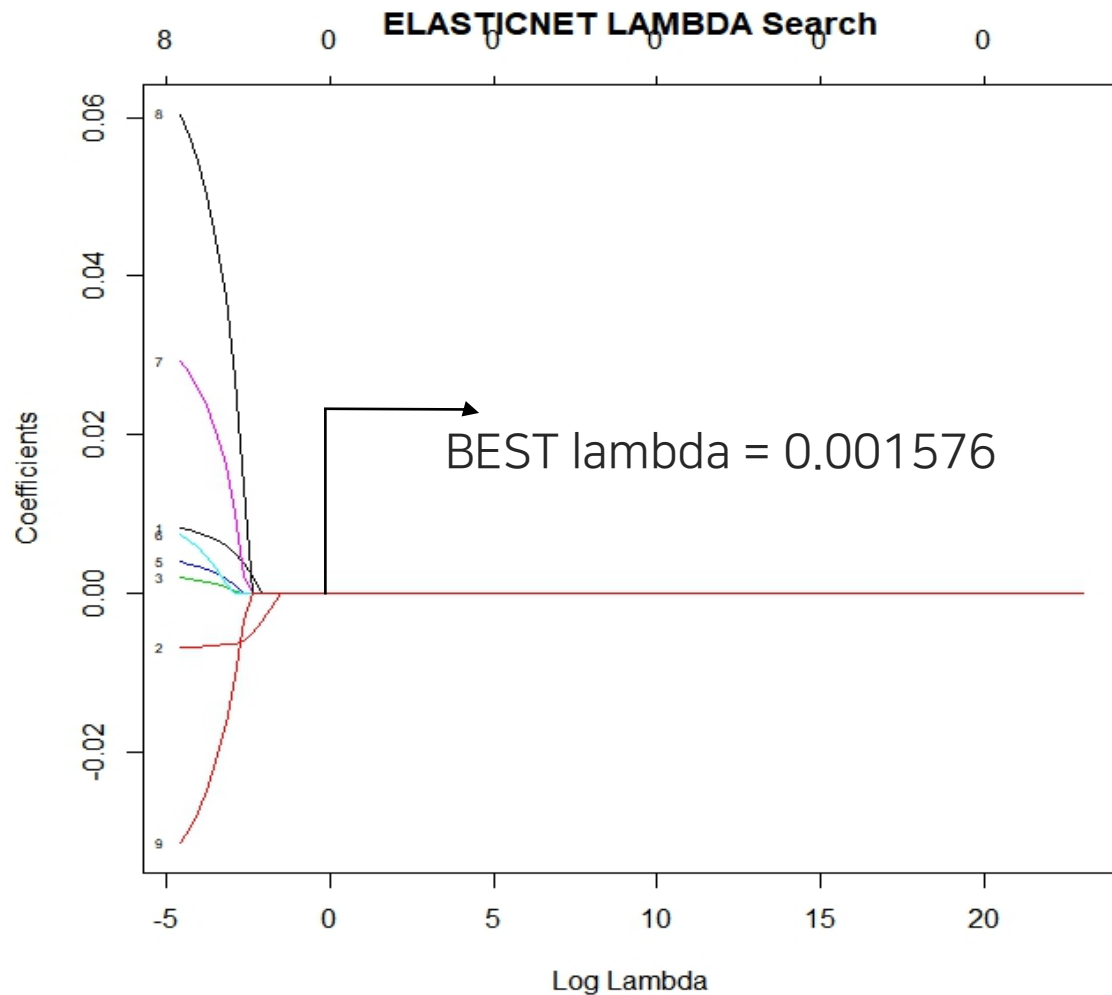
변수선택 시 로그 람다 값에 따른 평균 제곱오차

- 변수선택으로 변수를 제거할수록 평균 제곱오차가 증가함으로 변수선택시 변수 탈락 실행하지 않음



변수 개수에 차이에 따른 설명편차

- 변수가 8개 일 때에 80%의 설명력을 가지지만, 모든 변수를 가질 때에 최적의 설명력을 가짐
- 그래프가 급격히 꺾이지 않으므로 overfitting이 아니라고 판단



### 람다 값에 따른 변수들의 계수 추정치

- 가로 축은 벌점 모수에 자연로그를 취한 값
- 세로 축은 가로 축에 대한 회귀 계수 값
- 람다 값이 증가함에 따라 설명변수의 회귀계수가 0으로 축소되는 것을 보여줌
- 맨 윗부분 숫자는 각 벌점모수 값에 대한 설명변수의 개수를 나타냄

Model	Accuracy	Recall
Elastic net 회귀	0.6452	0.8965
변수	오즈비	
풍속	1.037695	
기압	0.961783	
습도	1.008236	
기온	0.995392	
수온	1.020383	
최대파고	1.043123	
유의파고	1.163329	
평균파고	1.350321	
파주기	0.852963	

훈련 데이터로 모델을 학습하고,  
'가공되지 않은 실제 사고데이터'를 예측한 결과

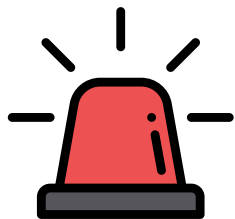
풍속과 습도, 수온, 최대파고, 유의파고  
평균파고는 모두 오즈비가 1을 초과했음  
이는 각 변수의 단위 값이 증가할 때마다  
사고 발생 예측(큰 값 1)으로 근접한다는 뜻  
정리하자면, 위의 변수들은 사고 발생과  
양의 발생 관계를 가짐

기압과 기온, 그리고 파주기는 오즈비가 0 이하로,  
단위가 증가할 수록 사고 미발생 예측(작은 값 0)  
으로 근접함정리하자면, 위의 변수는 사고 발생과  
음의 발생 관계를 가짐

04

## 솔루션 제안





## 위험지표 필요성

- 현재 기상에 대한 주의보로 호우주의보, 풍랑주의보 등이 존재
- 그러나 해양 상에서 선박 및 어선의 운항자에게 실질적으로 체감할 수 있는 직관적인 기상 위험도의 지표가 필요함
- 축적된 데이터와 사고 데이터를 이용하고, 모델의 정확도와 변수들의 오즈비를 고려하여 '사고발생확률'에 따른 위험지표를 개발하고자 함

## &lt;사고발생확률에 따른 위험지표 구간화&gt;

80% 이상	60% 이상	40% 이상	20% 이상	20% 미만
A 등급	B 등급	C 등급	D 등급	E 등급



### 해상사고 발생시

- 해상 사고는 육지로부터의 거리가 가깝지 않기에 대응이 느림
- 교통의 수단과 기상에 따라 효율적인 대처가 어려움



### 출항 보고

- 일반적인 선박의 경우에는 **이동 경로**를, 어선의 경우에는 어업 **장소**를 포함한 **시간대**를 출항 전에 인근 해양 경찰서 및 해양안전관리 기관에 보고
- 문제 발생시에 해상 사고에 대하여 **효율적이고 신속한 대처**를 도모
- 나아가서 추후의 해양 데이터 분석에 **기반의 되는 데이터**로 활용 가능



출항 전



출항 보고



문제 발생시 효율적 대응





## ① 메인 인터페이스 - 종합 정보 알림

- 사용자가 해역을 지정하면, 시간 단위의 기상을 확인 가능
- 축적된 데이터를 통해 위험등급을 알려주고, 사고 확률을 인지시킴으로써 경각심을 강조
- 사용자가 해역을 지정할 수 있도록 하여 유동적인 확인 가능

## ② 기상 기록실 메뉴

- 장소와 시간에 따른 기상을 확인할 수 있는 인터페이스
- 다른 장소와 과거의 시간까지도 기록이 되어 사용자로 하여금 기상에 대한 정보 파악을 용이하게 함
- 기상에 따른 사상자 발생 사례와 사고 기록을 볼 수 있음



### ③ 출항 보고

- 선박이나 어선의 출항시에 인근 해양 경찰소 및 해상안전관리 기관에 보고하는 프로세스
- 선박의 이동경로나 어선의 어업 장소, 시간을 전달 가능
- 위험등급 및 사고확률이 의무적으로 수행되길 권장
- 데이터의 축적으로 추후 선박 및 어선 분석에 활용
- 긍정적인 보상체계와 안정적인 제도로 일반화 중요

### ④ 긴급 연락망

- 선박이나 어선이 위험한 상황에 이르렀을 때, 혹은 도움이 필요한 순간에 긴급적으로 인근 해양 경찰소 및 해상안전관리 기관에 즉각 연결
- 전화연결 및 문자 발송이 가능하며, 선박에 대한 정보나 위치를 같이 전송하여 신속하고 효율적인 대처를 가능하게 함
- 연락망은 사용자가 임의로 추가 지정이 가능

05

## 기대 효과



기상 악화로 인한 사고 피해액 연 평균 약 240억원 절감 가능



기상 악화로 인한 사고 건수 연 평균 약 60건 예방 가능



사고 예방으로 인한 사상자 수 감소 가능



해양 사고에 대한 새로운 안전 체계 시스템 제안

# THANK YOU

---

Team  
S4