2023 날씨 빅데이터 콘테스트

해양안전분야

(기상에 따른 선박 닻 끌림 예측)

Team. 닻닻띠라라

김혜리 (kwc950515@naver.com)

임성수 (eric_best@naver.com)

정인영 (wjddlsdud033@naver.com)

Index

1 분석 주제 이해

주제 요약 및 관련 이슈

분석 목표

2017~2021년 부산항 사고발생 현황

(자료: 부산항운노조)

연도	경상	중상	사망
2017년	69건	47건	0건
2018년	82건	59건	5건
2019년	93건	31건	3건
2020년	113건	29건	3건
2021년	122건	21건	1건
합계	479건	187건	12건

2011~2020년 전국 항만사업장 산업재해

연도	명	연도	명
2011년	334명	2016년	242명
2012년	327명	2017년	220명
2013년	292명	2018년	268명
2014년	292명	2019년	274명
2015년	273명	2020년	278명
		합계	2800명

(자료 : 해양수산부)

"없어지지 않는 주요항의 선박 닻끌림 사고 "

화물선 닻 끊고 600m 떠밀어 좌초시킨 강풍의 위력

'오션탱고호' 유출 기름 영도 · 부산대교까지 확산

부산항 묘박지 해상 선박끼리 부딪혀…평형수 일부 유출

기상청, 주요항만 정박지 사고 예방 위해 맞춤 해양기상정보 제공

해상 교통관제 시스템, 선박사고율 크게 낮춰

[10] 김혜리, 2023-07-24

DATA

- busan_drag(닻끌림)
- ulsan_drag(닻끌림)
- 기상 데이터
  해양 데이터

2021.01~2022.06

분석 과제

‘기상 상태에 따른 닻 끌림 발생 여부 분석’

‘현재 시점에서 30분~1시간 후 닻끌림 발생 예측 모델 아이디어 제안‘

‘향후 닻끌림 자동시스템에 예측모델로 활용’

2022.07~2023.03

기상 및 해양데이터를 통한 분석을 진행하면서, 선박의 닻끌림 여부를 예측하고,
주요 항만 정박지의 맞춤 해양정보와 함께 이용하여 정박지 사고 예방효과 기대

2 데이터 이해 및 탐색

Q. 주어진 데이터 셋의 구조는?

[ER Diagram]

busan / ulsan (anch / drag)	Answer (busan / ulsan) 닻 끌림 발생 정보	기상데이터 (KMA)	해양데이터 (KHOA, KHNP)
Num (선박번호, PK1)	NUM (임의 선박번호)	YYMMDDHHMI (시간)	YYMMDDHHMI (시간)
Time (시간, PK2)	area (발생장소)	STN/STN_NAME	STN_NAME
Latitude	year (년도)	WD (풍향)	WS (유속)
Longitude	mon (월)	WS (풍속)	WD_point (유향)
SOG (대지속력)	day (일)	max_wh (최대파고)	WD (유향, degree)
COG (실침로)	hour (시간)	sig_wh (유의파고)	
HDG (선수미선)	Min (분)	mean_wh (평균파고)	
	lat (닻끌림 위도)		
	lon (닻끌림 경도		

[분석 Insight]

Answer데이터를 이용하여 busan/ulsan 데이터에 target을 생성하고
지점(명)과 시간을 기준으로 기상 및 해양데이터를 사용하자.

Q. 데이터 내 닻끌림 발생은?

< busan answer date 中 1번선박 >

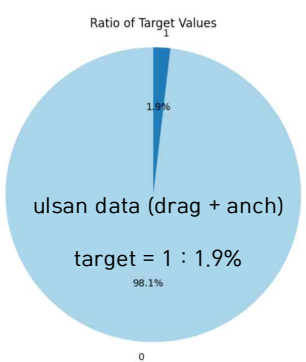
	busan_answer.area	busan_answer.year	busan_answer.num	busan_answer.mon	busan_answer.day	busan_answer.hour	busan_answer.min	busan_answer.lat	busan_answer.lon
1	BUSAN	2021	1	1	17	2	4	N35.045368	E129.071012

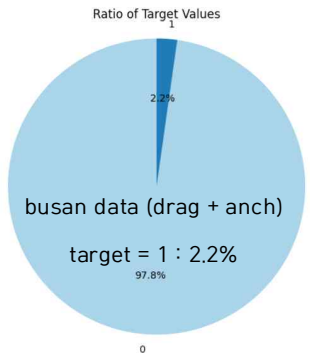
< busan drag >

	busan_drag_train_final.num	busan_drag_train_final.time	target
570	1	2021-01-17 01:34:00	1
571	1	2021-01-17 01:46:00	1
572	1	2021-01-17 01:49:00	1
573	1	2021-01-17 01:52:00	1
574	1	2021-01-17 01:58:00	1
575	1	2021-01-17 02:01:00	1
576	1	2021-01-17 02:04:00	1
577	1	2021-01-17 02:07:00	1
578	1	2021-01-17 02:13:00	1
579	1	2021-01-17 02:19:00	1
580	1	2021-01-17 02:22:00	1
581	1	2021-01-17 02:28:00	1

[-30분]

[+30분]





Answer 데이터의 시간과 경도/위도를 기준으로 busan/ulsan drag 데이터셋에 target=1을 넣었다.

좀 더 예민한 예측모델 구축을 위해 target=1이 찍힌 지점의 ±30분을 닻끌림이 발생하였다고 판단하였다.

[분석 Insight] 데이터 불균형 문제가 존재한다고 판단하였다.

Q. 선박별 기상 및 해양데이터 이용방안은?

[기상 및 해양데이터 탐색]

구분	지역	지점	관측장비	요소
기상관측자료 (기상청)	울산	이덕서	등표	풍향·풍속 (1분 자료)
		간절곶	파고부이	파고 (유의,최대,평균 /시간자료)
		당사		
	부산	오륙도	등표	풍향·풍속 (1분 자료)
		장안	파고부이	파고 (유의,최대,평균 /시간자료)
		기장		
		오륙도		
		다대포		
해양부이관측자료 (한수원, 조사원)	울산	한수원_고리	부이	유향·유속 (1분 자료)
	부산	조사원_송정	부이	유향·유속 (1분 5분 자료)
		조사원_해운대		

파고부이와 부이와 같은 경우 관측지점이 다양함을 확인

기상 데이터마다 시간 단위가 다름을 확인

=> 시간 단위를 맞추어 알맞은 기상 및 해양데이터 활용 필요

기상 및 해양데이터의 모든 변수들을 닻끌림 데이터의 파생변수로 두고 분석에 활용하기 어렵다고 판단하였다.

[분석 Insight] 각 선박별 위치(위도,경도) 기반으로 가장 가까운 기상 및 해양데이터 정보를 활용하자

< 울산항 정박지 정보 >

정박지	시설코드	위 치	비고 (GT)
M1	WAM-01	상기 A, B, P, Q의 4지점을 연결하는 선내의 해면	
M2	WAM-02	상기 B, C, D, S, R의 5지점을 연결하는 선내의 해면	
M3	WAM-03	상기 N, O, P, R, S의 5지점을 연결하는 선내의 해면	2천톤 이하
M4	WAM-04	상기 D, E, M, N의 4지점을 연결하는 선내의 해면	
M5	WAM-05	상기 E, F, K, L, M의 5지점을 연결하는 선내의 해면	
M6	WAM-06	상기 F, G, J, K의 4지점을 연결하는 선내의 해면	
M7	WAM-07	상기 G, H, I, J의 4지점을 연결하는 선내의 해면	
E1	WAE-01	① N 35° 27′ 59.0″, E 129° 24′ 51.4″ ② N 35° 27′ 59.0″, E 129° 25′ 34.7″ ③ N 35° 26′ 46.7″, E 129° 27′ 49.3″ ④ N 35° 26′ 13.6″, E 129° 24′ 39.5″ ⑤ N 35° 27′ 43.4″, E 129° 24′ 04.7″	1만톤 이하
E2	WAE-02	① N 35° 26′ 13.6″, E 129° 24′ 39.5″ ② N 35° 26′ 46.7″, E 129° 27′ 49.3″ ③ N 35° 25′ 29.8″, E 129° 28′ 25.9″ ④ N 35° 25′ 12.7″, E 129° 25′ 03.1″	3만톤 이하
E3	WAE-03	① N 35° 25′ 12.7″, E 129° 25′ 03.1″ ② N 35° 25′ 29.8″, E 129° 28′ 25.9″ ③ N 35° 23′ 02.5″, E 129° 27′ 26.4″ ④ N 35° 24′ 11.0″, E 129° 25′ 27.0″	2만톤이상
W1	WAW-01	N 35°27′17.0″, E 129°23′23.0″ 중심의 반경 400m 원내의 해면	2만톤 이하
T1	WAT-01	N 35°30′37.3″, E 129°27′17.7″ 중심의 반경 300m 원내의 해면	5천톤 이하
T2	WAT-02	N 35°30′57.0″, E 129°27′17.7″ 중심의 반경 300m 원내의 해면	2만톤초과 ~ 5만톤이하
T3	WAT-03	N 35°31′40.2″, E 129°27′34.0″ 중심의 반경 250m 원내의 해면	2천톤이하

< 울산항 정박지 별 평균선회반경 >

울산항 정박지			
정박지	평균 선회반경(m)	정박지	평균 선회반경(m)
E1	290	M1	90
E2	400	M2	90
E3	483	M3	90
W1	400	M4	90
T1	290	M5	90
T2	95	M6	90
T3	95	M7	90

- 정박지 별 평균 선회반경과, 정박지내 선박의 시설능력(G/T)가 다른것을 확인,
  - 이는 외력의 작용으로부터 닻끌림이 인지되는 주요 변수일것이라 판단.

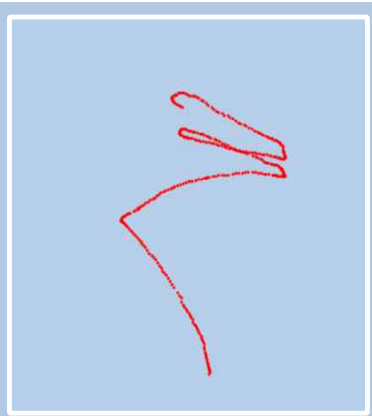
[분석 insight] 각 선박별 정박지를 생성하여 분석에 이용하자.

< 닻끌림 선박의 항적 패턴 >

< train 데이터셋에서 닻끌림이 발생한 선박 시각화 >

< 일(-) 방향 패턴 >

ulsan train 데이터의 69번 선박



< 지그재그 패턴 >

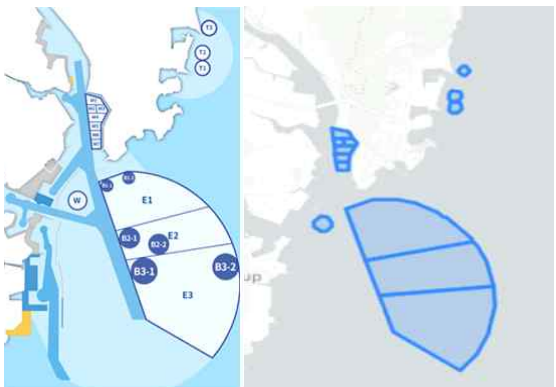
ulsan train 데이터의 38번 선박

‘닻끌림’이 발생한 선박의 데이터를 시각화한 결과,

정박선의 항적이 선회반경을 넘어 일자나 지그재그를 그리고 있는 것을 확인하였다.

3 데이터 전처리

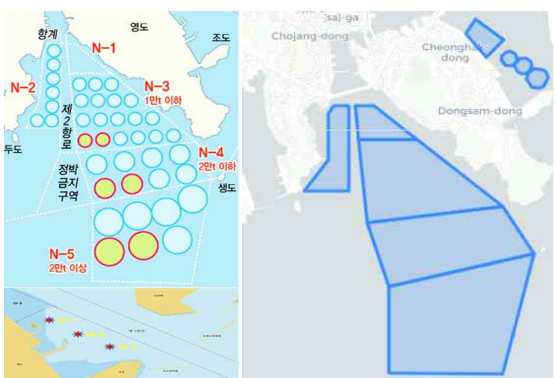
울산

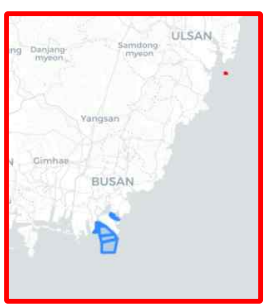


- EPSG:5186좌표계와 EPSG:4326좌표계를 이용해 정박지를 형성
- 부산의 경우 추가적인 정박지 o2를 생성
- 부산항의 M7,M8,M9 정박지는 [redacted]학반경을 고려한 원형 정박지로

선박의 위도와 경도에 따른 위치가 해당되는 정박지가 두개 이상일경우, 처음 배가 있었던 위치를 기반해 정박지 선정

부산





busan_anch 데이터의 112번 선박의 위치가 울산쪽에 위치해

정박지 정보를 ‘outlier’로 두었다

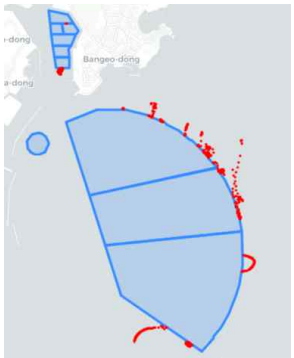
슬라이드 13 페이지

[10] 정인영, 2023-07-31

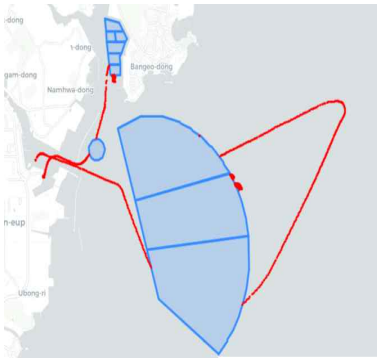
이에 해당의 정박지 관련 내용 쓰고

아래 내용 M선박지로 합친 이유

파생변수인 정박지 컬럼은 선박의 위치를 나타내면서 선박의 무게 대상 내에 해당하는 중요한 변수이므로 유사한 M1-M7 부산의 M8,M9 경우 선박의 무게가 동일하며 정박지 간 위치차이가 크지 않기에 각각 하나의 정박지로 구분하여 주었습니다.



ulsan_drag



ulsan_anch

busan_drag

busan_anch

ulsan_drag, ulsan_anch, busan_drag, busan_anch 데이터 중정박지에 위치하지 않은 선박들을 시각화 한 결과

운항중인 선박 또는 닻끌림이 일어난 선박이 있음

정박지 밖 데이터가 닻끌림 발생인 데이터로구분될 확률이 높으므로 유의하여 모델링 참고

선박 데이터

- S0G (대지속력)

102.3 값을 갖는 COG. HDG값이 모두 null 값이었음을 확인하였고 이상치라고 판단하여 null값으로 두고 결측치 처리 진행

- COG (실침로)

선박이 이동하는 방향의 방위각으로 0°=360°로 두고 결측치 처리 진행

- HDG (선수미선)

선박이 이동하는 방향의 방위각으로 0°=360°로 두고 결측치 처리 진행

KNN-Imputer

각 표본의 결측값은 학습 셋에서 찾은 n_neighbors 가장 가까운 이웃의 평균값으로 사용하여 대체된다.

즉. 원하는 인접 이웃 수의 가중 또는 가중 평균을 사용하여 결측값을 대체 하는 방식이다.

결측값 처리

각 선박당 가장 가까운 이웃데이터들을 통하여 결측치 처리 [sog] [cog] [hdg]

*설명자료에 나온 결측치들은 모두 nan(null) 값으로 변환한뒤 전처리를 진행하였습니다.

설명 변수의 결측치 주위에 가까운 몇 개의 데이터를 통해 결측치를 대체하는 방법은 KNNImputer를 사용하여 결측치를 대체하였습니다.

기상청 등표 (KMA_LightBeacon)

- 컬럼값이 모두 같은 행들이 존재하는 것을 확인 후 중복제거
- 울산 (이덕서) / 부산 (오륙도)
- 1분 간격의 시계열 데이터 울산 및 부산 데이터 시간대에 존재하지 않는
- 시간대의 풍향,풍속 컬럼은 선형보간법을 사용하여 처리함

기상청 파고부이 (KMA_PaogBuoy)

- 울산 (간절곶.당사) / 부산 (장안, 기장,오륙도,다대포)
- 각각의 관측점에 따라 전처리 진행
- 1시간 간격의 시계열 데이터. 부산/울산 데이터에 결합하였을 때
- 발생하는 풍향, 풍속컬럼의 결측치들은 선형보간법을 통해 처리함

슬라이드 16 페이지

[31] 정인영, 2023-07-31

기상레이더의 대한 전처리과정을 설명드리겠습니다.

기상청 등표의 경우 울산의 이덕서와 부산의 오륙도 두 관측지점으로 [illegible] 있으며 선형보간법을 통해 풍향,풍속 결럼 결측치를 처리하였습니다.

기상청 파고부이는 울산의 간절곶 당사 부산의 장안 기장 오륙도 다대 [illegible] 어준 후 선형보간법을 통해 풍향,풍속 결럼 결측치를 처리하였습니다

한수원_파고부이 (KHNP_Buoy)

울산 (고리)

1분 간격으로 시계열 구성. 결측치들은 선형보간법을 통해 처리

해양조사원_파고부이 (KHOA_Buoy)

부산 (송정 : 5분간격) / 부산 (해운대 : 1분 간격)

5분 간격으로 이루어진 송정을 1분 간격으로 데이터 확장 후

결측치들을 선형보간법을 통해 처리함

김정인영 2023-07-31

한국해양과학기술원은 울산의 고리 1개의 관측지점으로 구성되어 있으며 부분 관측 오류 없이 회귀대의 데이터가 존재하여 선형보간법을 통해 결측치를 처리하였습니다.

해양조사원 관고부이는 부산의 송정과 해운대 등 관측지점으로 구성되어 있으며 부분 관측 오류 없이 회귀대의 데이터가 존재하여 선형보간법을 통해 결측치를 처리하였습니다.

해양조사원_파고부이 <wd_point 범주형 변수 존재>

1. wd_point의 value_counts확인

2. 각 범주별 wd변수의 평균값 확인

3. wd_point 결측치 행의 wd값을 구한 평균값과 비교해 근사치 값 wd_point 값으로 대체

wd_point의 value		Value별 wd변수 평균값	wd_point의 value		Value별 wd변수 평균값
E	79997	[88.95279822993362	NNE	7469	22.9674655241266556
W	45167	267.4902251643899	SW	6452	226.42978921264725
WSW	34177	251.26693390291717	NW	6127	313.6461563571079
ENE	24094	70.10081348053457	N	5114	159.53656628861947
ESE	23020	110.60742832319723	NNW	4448	337.36668165467626
WNW	13331	290.6028054909609	SSE	3860	156.36424870466323
NE	9375	45.62325333333333	SSW	3731	203.68319485392655
SE	9051	133.60844105623687	S	3406	180.0971814445097

슬라이드 18 페이지

[5] 정의영, 2023-07-31

해양조사원 파고 부이 wd_point라는 범주형 변수에 대한 결측치는 3가지 과정을 통해 진행하였습니다.

먼저 wd_point 중 결측치가 아닌 값의 value_counts를 확인한 뒤 수치형 변수인 wd의 통계량을 통해 wd_point별 wd의 평균값을 확인하고 wd_point의 결측치가 존재하는 행에 대하여 wd 값을 이용하여 wd_point별 평균값과 가장 가까운 값으로 결측치를 대체하였습니다.

기상 데이터 합치기 / 변수선택

	busan_drag_train_final.num	busan_drag_train_final.latitude	busan_drag_train_final.longitude	kma_pagobuoy_train.stn_name
466	1	"N35.049533"	"E129.074067"	오륙도
535	1	"N35.048995"	"E129.073660"	오륙도
696	3	"N35.006715"	"E129.063955"	다대포
939	4	"N35.019232"	"E129.044707"	다대포
1055	4	"N35.016032"	"E129.041622"	다대포

선박데이터의 경도,위도와 기상데이터의 관측지점의 경도,위도를 비교하여 선박과 가장 가까운 관측지점의 기상데이터와 결합

울산

sog	max_wh	ligthbecon_wd	khnp_bouy_ws
cog	sig_wh	lightbecon_ws	stn_num
hdg	mean_wh	khnp_bouy_wd	

부산

sog	max_wh	ligthbecon_wd	khnp_bouy_wd_point
cog	sig_wh	lightbecon_ws	khoa_bouy_wd
hdg	mean_wh	khnp_bouy_ws	stn_num

[12] 정인영, 2023-07-31

선박데이터의 경도, 위도와 각각의 기상데이터 관측지점의 경도, 위도를 비교하여 선박의 위치와 가장 가까운 관측지점의 기상데이터를 결합하였습니다. 예시를 보면 경도 위도에 따라 오륙도, 다대포 중 더 가까운 관측지점의 기상데이터로 결합된 것을 확인할 수 있습니다. 가까운 관측지점 기반으로 기상데이터와 선박데이터 결합한 결과 울산 11개, 부산 12개의 컬럼으로 최종 데이터프레임을 구성하였습니다.

4

모델링

1. 데이터 불균형

-> Oversampling

2. 범주형 변수처리

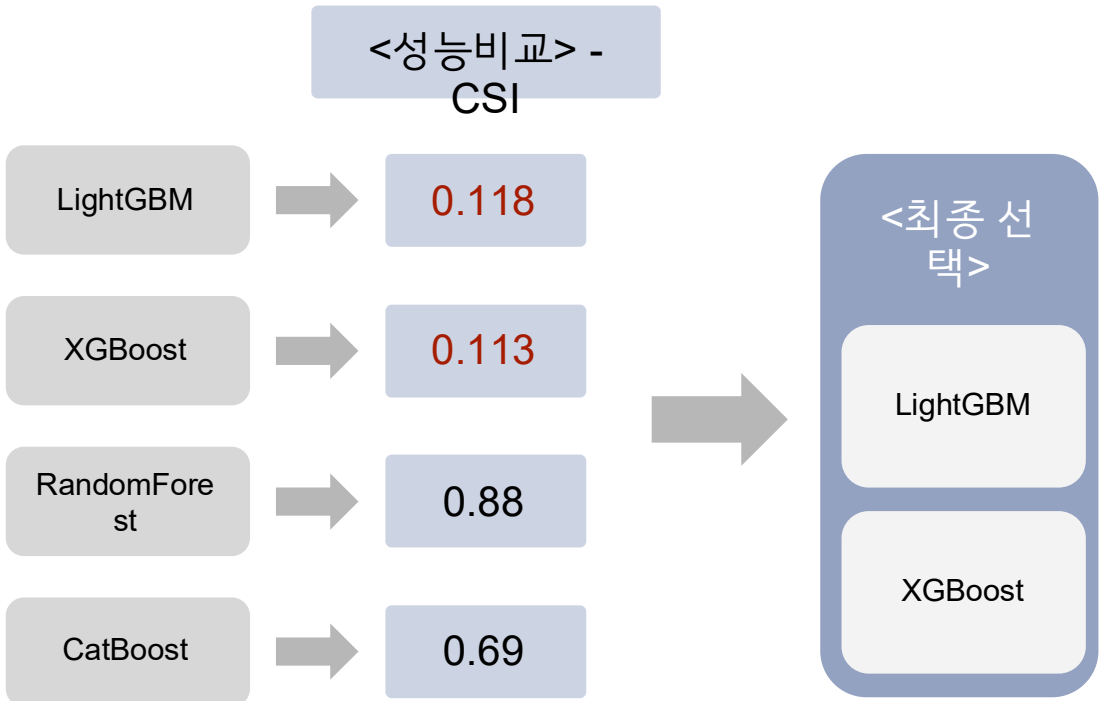
-> Label Encoding

3. 검증

-> Stratified-KFold

4. 하이퍼파라미터 최적화

-> Optuna



슬라이드 21 페이지
71전인영, 2023-07-31
다음은 범주형 데이터가 2%정도로 데이터 불균형이 심하여 oversampli
ng 중 smote 기법으로 샘플링을 하여 불균형을 해결하였고 Stratified K-Fold
를 진행, 하이퍼파라미터 최적화에는 Optuna를 사용하였습니다.
LightGBM XGBoost RandomForeset CatBoost 4 종류의 모델을 이
와 같은 방법을 통해 모델링하여 CSI를 확인한 결과 보이는 것과 같이
LightGBM과 XGBoost의 결과가 상대적으로 좋았습니다. 이에 가장
은 두 모델을 최종 모델로 선정하였습니다.

< 최적의 하이퍼파라미터 >

LightGBM

-ulsan-
max_depth : 13
learning rate : 0.065
n_estimators : 1962
num_leaves : 371
colsample_bytree : 0.61
subsample : 0.6
reg_alpha : 0.364
reg_lambda : 5

-busan-
max_depth : -1
learning rate : 0.085
n_estimators : 1595
num_leaves : 31
colsample_bytree : 0.83
subsample : 0.76
reg_alpha : 1.46
reg_lambda : 1.5

XGBoost

-ulsan-
max_depth : 7
learning rate : 0.75
n_estimators : 6000
eta : 0.07
reg_alpha : 24
reg_lambda : 98
min_child_weight : 22
colsample_bytree : 0.75

-busan-
max_depth : 9
learning rate : 0.8
n_estimators : 6000
eta : 0.04
reg_alpha : 6
reg_lambda : 86
min_child_weight : 16
colsample_bytree : 0.94

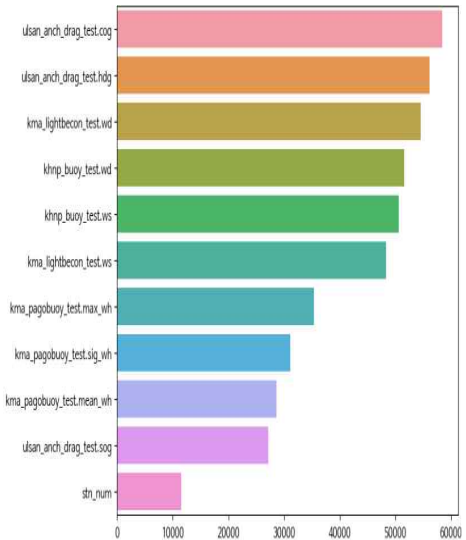
→ **Hard Voting** → **0.14**

[8] 정의영, 2023-07-31

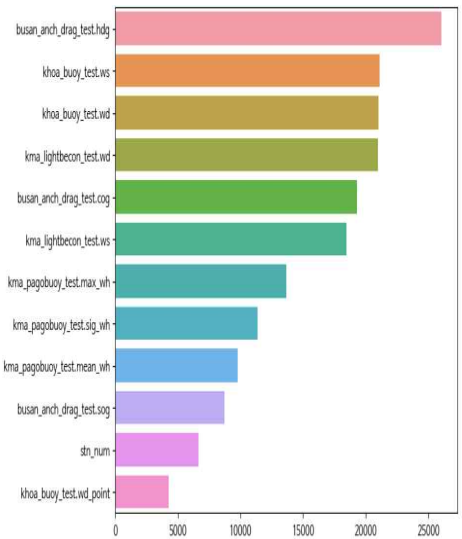
[8] 모델들의 파라미터를 조절하여 얻은 예측값들을 하드 보팅을 통하여 CSI 지수를 0.14까지 끌어올렸습니다

LightGBM

울산

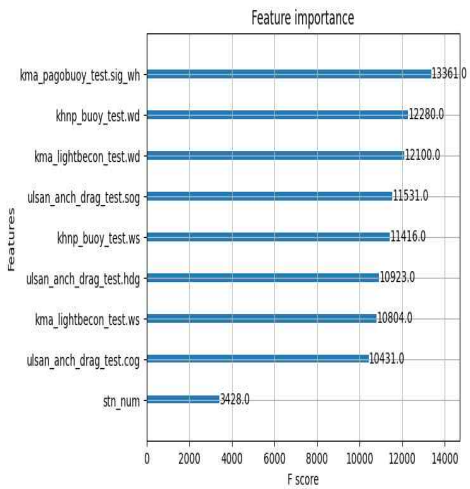


부산

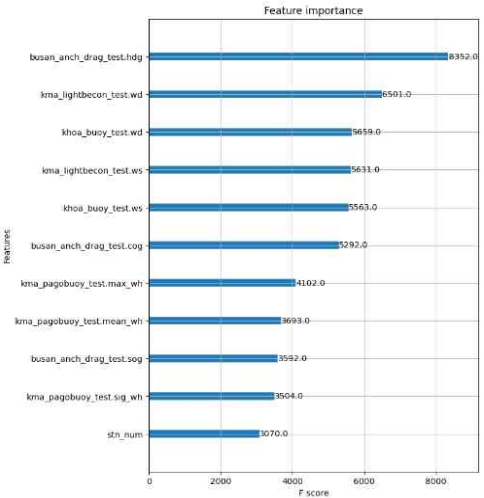


XGBoost

울산



부산

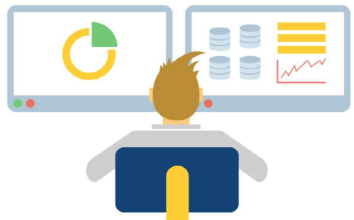


5 활용방안 및 기대효과

활용 방안 | 기대 효과

기상청

해양경찰청
KOREA COAST GUARD



예측된 기상정보, 해양정보, 닻끌림 예측모델에 기반하여 나눈 위험군을 4단계를 통해

닻끌림 발생위험 시 관련 종사자들을 대상으로 한 비상대응 매뉴얼 제시

이미지 출처: Eucalyp,산림청, https://www.pngwing.com/ko/free-png-ssmng

슬라이드 25 페이지

[01 정인영 2023-07-31

기상청과 해양경찰청에서 예측된 기상정보와 해양정보 및 닻끌림 예측 모델을 기반으로 닻끌림이 예측되는 선박 간의 사고를 예방하기 위해 위험군을 4가지 단계로 나누어 관련 종사자들을 대상으로 비상 대응 매뉴얼 제시

발생가능성의 확률에 따라 임계값을 통해
0.5~0.6(관심)
0.6~0.7(주의)
0.7~0.8(경계)
0.8~1(심각)으로 나누어 사전에 종사자들에게 알려주어 위험도에 맞게 선박의 위치를 조정한다면 선박간의 사고를 예방하는데 큰 도움을 줄 것이라 생각합니다.

#안전

- 데이터 분석으로 현재시점의 1시간 후 예측시점을 정확히 인지하고.
- 울산.부산항에 정박하는 선박에게 안전하고 정확한 닻끌림 예측서비스를 제공하여
- 기상 악화 시 발생하는 닻끌림 및 해양사고를 예방할 수 있다.

#효율

- 올바른 정박과 항해시 기상청에서 제공하는 정박지 맞춤형 해양기상정보를 활용해
- 선박교통 및 항만 입출항에 도움을 주어 해상교통의 효율상승을 도모할 것이다.

#환경

- 위 데이터 분석과정과 결과를 활용하여 사고를 미연에 예방하여
- 기름유출 및 해양쓰레기들의 발생빈도를 줄여 해양오염사고의 피해를 줄일 수 있을것으로 기대됨.