

# 연안 사고 예방 아이디어 제안서

함형빈, 김연우

## 요약

본 제안서는 연안 사고 예방 아이디어를 구하기 위해 “해양경찰청\_연안 사고 이력” 데이터를 기준으로 탐색적 데이터 분석(EDA)을 수행했다. 해당 자료는 공공데이터 포털에서 수집했다. EDA를 통한 각 사고 요인에 대한 빈도수와 시각화 그에 따른 필요한 사고 예방 아이디어를 도출했다. 이후 앙상블 모델을 분류 모델로 사용하여 각 요인에 대한 사고 유형 분류를 시도했다. KNN, Decision Tree, Random Forest, LightGBM 4가지 앙상블 모델을 사용했고 시간 내에 시도한 경우 중 가장 성능이 높았던 Random Forest로 정했다. 결과적으로 분류 정확도는 70%로 높진 않았다. 데이터의 양과 데이터 전처리가 잘 됐었다면 개선이 가능할 것 같다. 연안 사고 예방 아이디어는 크게 3가지로 도출했다. EDA를 통해 파악된 파출소 업무 분담 필요성, 각 파출소별 사고 유형별 예측 시스템, 사고 발생 원인별 대체 방안 아이디어 제안을 도출했다.

키워드 - 딥러닝, 파출소 업무 분담, 파출소별 사고 유형 확률 예측, 안전수칙 준수 인증 시스템, 수위 감지 시스템

## I. 데이터 소개 및 전처리

공공데이터 포털의 “해양경찰청\_연안 사고 이력”, “해양경찰청\_파출소 위치” 데이터를 CSV 파일과 오픈 API로 활용했다. 첫 번째 데이터는 2017~2020년 연안 사고 이력에 관한 데이터로 발생지역, 관할서, 발생시각, 사고 유형, 사고 원인 현황을 정리한 데이터이고 두 번째 데이터는 각 해양 파출소별 이름, 위도, 경도 데이터이다.

데이터 전처리는 각 컬럼에 있는 결측값을 모두 ‘무지역’, ‘무관할’ 등 값을 대체했다. 위도\_도, 위도\_분 등 각 세분된 위·경도 값을 위도, 경도로 계산해 컬럼으로 자료에 추가했다. 그리고 발생 일자를 연도, 월로 나눴고 발생시각 시간대가 오전인지 오후인지, 계절이 언제인지 파악했다. 또한 광역시도에 따라 동해, 서해, 남해, 제주로 나눠 추가했다. 서해안은 충청남도, 인천광역시, 전라북도, 경기도를 포함했고, 남해안은 전라남도, 경상남도, 부산광역시, 동해안은 강원도, 경상북도, 울산광역시, 제주는 제주특별자치도, 제주특별자치시를 포함했다.

앙상블 모델을 사용하기 위해 자료를 다른 변수로 복사했고 ‘광역시도’, ‘발생지역’, ‘해경서’, ‘파출소’, ‘발생요일’, ‘사고유형\_대분류’, ‘사고유형\_

세분류’, ‘장소’, ‘사고원인’, ‘위험구역지정’, ‘시간대’, ‘계절’, ‘해안’을 제외한 모든 칼럼은 제거하고 종속 변수와 독립변수를 나눴다. 이후 Train, Test 데이터 셋을 7:3 비율로 나눴다. 시계열성을 없애기 위해 shuffle=True로 지정했고 반복적인 수행에도 같은 결과가 나오도록 random\_state=10으로 정했다. Train, Test의 독립 데이터는 모두 원-핫-인코딩으로 dummy(더미) 화를 했다. 종속 데이터는 {"추락익수":1, "고립":2, "기타익수":3, "추락":4, "표류":5, "수상레저익수":6, "기타":7, "수중레저익수":8, "고립익수":9, "수중산업익수":10, "수상산업익수":11}으로 라벨링했다.

라벨링된 종속 데이터의 비율

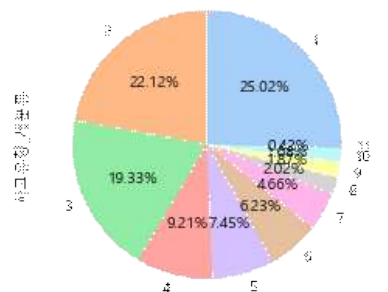


그림1. 라벨링 된 종속 데이터의 비율

## II. 탐색적 데이터 분석(EDA)

(그림 2) 광역시도 별 사고 발생 수 비율은 충청

남도, 인천광역시, 전라남도, 제주, 경상남도, 부산광역시, 강원도, 경상북도, 전라북도, 경기도, 울산광역시 순으로 많았다.

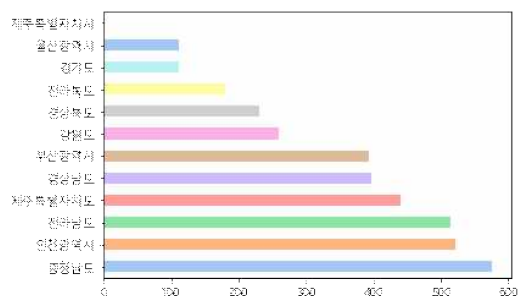


그림2. 광역시도별 사고 발생 빈도수

(그림 3) 하늘바다 파출소가 다른 파출소보다 가장 많은 출동량을 보여주고 제주, 무관할, 마검포 순으로 출동량이 많다. 하늘바다 파출소(인천)는 다른 파출소에 비해 훨씬 많은 출동을 하고 있다. 이는 파출소 과부하로 이어져 출동에 문제가 생길 수 있다. 이런 상황에 해당하는 지역의 파출소 업무 분담 문제점을 해결해야 할 것 같다.

강원도는 [속초, 강릉, 삼척] 파출소, 경기도는 [대부, 안산, 신항만], 경상남도는 [통영, 장승포, 마산]. 경상북도는 [영일만, 구룡포, 감포], 부산광역시는 [남항, 광안리, 기장], 울산광역시는 [방어진, 진하, 울산항], 인천광역시는 [하늘바다, 영흥, 인항], 전라남도는 [봉산, 북항, 서산], 전라북도는 [변산, 새만금, 비응], 제주도는 [제주, 서귀포, 한림], 충청남도는 [마검포, 학암포, 모항] 파출소가 Top 3 출동량을 가진다.

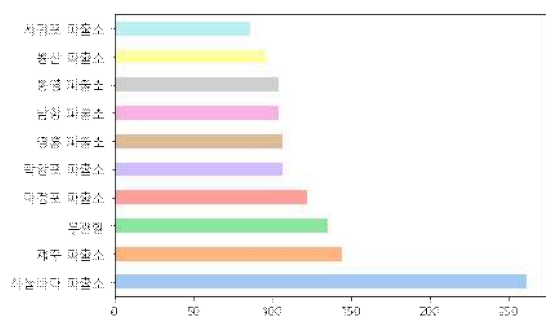


그림3. 파출소별 출동 빈도수

(그림 4) 충청남도의 파출소는 균등하게 출동하는 편이다. 하지만 인천광역시의 신항만 파출소는 12건으로 다른 파출소에 비해 현저히 적고 하늘바다 파출소의 출동량은 충청남도 최고 출동량인 마검포 파출소의 2배 이상이다. 이는 신항만 파출소가 경기도를 담당하고 있는 비율이 존재하기 때문

에 인천으로 출동하는 비율이 적은 것 같다.

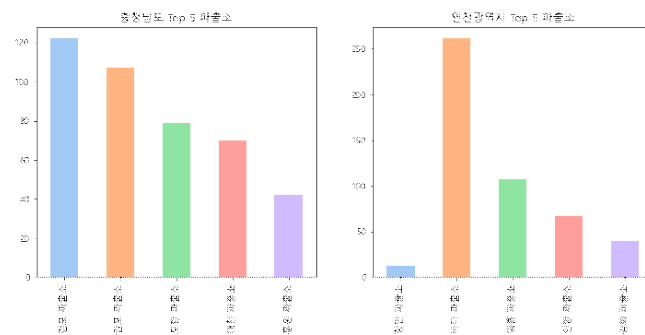


그림 4. 충청남도, 인천광역시 파출소 출동량 비교 (그림 5) 추락익수(24.25%), 고립(22.67%), 기타 익수(19.79%), 추락(9.49%), 표류, 수상레저익수, 기타, 수상레저익수, 고립익수 등(약 25%)의 유형으로 사고가 발생한다. (그림 6) 상위 4개의 사고 유형별 사고원인을 보면 부주의, 자살, 조석 미인지, 음주가 대표적으로 차지했다. 사고원인에 대한 대책이 필요할 것 같다.

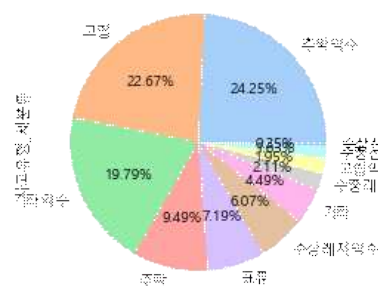


그림5. 사고 유형 비율

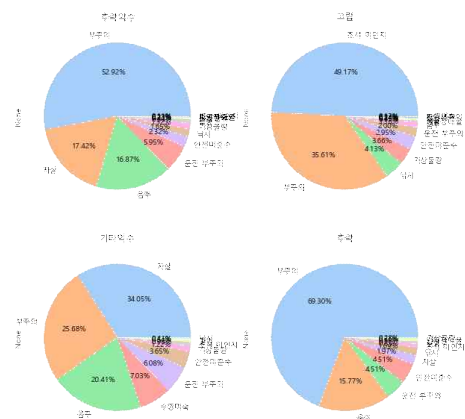


그림6. 상위 4개 사고유형별 사고 원인 비율 (그림 7) 4계절 사고 비율을 보면 여름(39.73%), 가을(26.82%), 봄(19.81%), 겨울(13.64%) 순으로 사고가 일어난다. 여름엔 해수욕을 이용하는 여행객의 사고로 늘어난 것으로 예상되고 겨울엔 추위로 인해 바다를 찾는 사람이 상대적으로 적어 사고

비율이 적은 것 같다.



그림 7. 계절별 사고 비율

동해, 서해, 남해, 제주의 사고 비율은 서해(37%), 남해(35%), 동해(16%), 제주(12%) 비율로 사고가 발생한다. (그림 8) 해안별 사고유형을 보면 남해안은 추락익수 사고가 부주의로 인한 원인으로 가장 많이 일어났고 그에 절반 비율로 각 기타익수 사고, 추락익수 사고가 자살의 원인으로 사고가 일어났다. 동해안은 추락 사고가 부주의 원인으로 가장 많았고 추락익수, 표류 사고가 부주의 원인으로 많았다. 서해안은 고립의 사고가 조석 미인지, 부주의 원인으로 전체 서해 사고의 50%로 큰 비율을 차지했다. 제주는 추락익수, 기타익수, 추락 사고가 부주의 원인으로 가장 많이 발생했다.

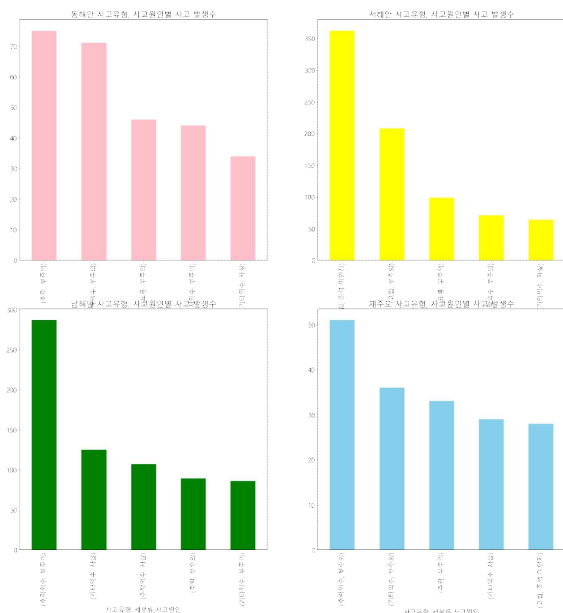


그림 8. 해안별 사고유형, 사고원인별 빈도수

### Ⅲ. 모델링 (사고 유형 예측 시스템의 예시)

데이터 전처리에서 데이터를 복사 후 7:3 비율로 나누고 독립 데이터는 dummy 화 종속 데이터는 라벨링을 했다.

남	여
흡연	비흡연
비흡연	흡연

남_흡연	남_비흡연	여_흡연	여_비흡연
1	0	0	1
0	1	1	0

표 1. 더미화 예시

(표 2) KNN 모델을 사용했을 때 Train 데이터에 대한 정확도는 71%, Test 데이터에 대한 정확도는 59%로 나왔다. Decision Tree 모델의 Train 정확도는 80%, Test 정확도는 68%로 상승했다. Random Forest 모델의 Train 정확도는 82%, Test 정확도는 72%로 나왔다. LightGBM 모델의 Train 정확도는 89%, Test 정확도는 73%로 나왔다. LightGBM의 경우 Random Forest 모델보다 정확도가 높지만, 과적합 가능성이 높다. 따라서 Random Forest 모델로 정했다. 해당 분류 모델은 데이터의 수도 적을 뿐만 아니라 전처리 또한 완벽하지 않다. 하지만 예측 모델을 참고해 순찰을 더 효율적으로 할 수 있을 것이라 예상된다.

	Train_Set_F1_Score	Test_Set_F1_score
KNN	71%	59%
Decision Tree	80%	68%
Random Forest	82%	72%
Light GBM	89%	73%

표 2. 모델별 정확도

## Ⅳ. 아이디어 제안

### 1. 파출소 업무 분담 및 조정

EDA에서 확인한 신항만 파출소는 인천광역시와 경기도를 같이 출동하고 있으며 다른 파출소보다 출동량이 많이 적다. 하늘바다 파출소는 인천 다른 파출소별 출동량보다 2배 이상의 출동량을 가져가고 있다. 충청남도도 인천광역시와 사건이 비슷하게 발생했지만, 파출소별 출동량이 비슷하다.

하늘바다, 신항만 파출소처럼 출동량 차이를 크게 가지지 않기 위해 위치나 업무분담을 하도록 한다. 예로 하늘바다 파출소의 출동량을 분산시키기 위해 상대적으로 출동량이 적은 신항만 파출소의 위치 또는 출동 지역을 하늘바다 파출소와 고려해 조정한다. 이런 파출소 업무분담 및 조정은 효율적인 출동과 연안 사고 대처에 도움이 될 것이다.

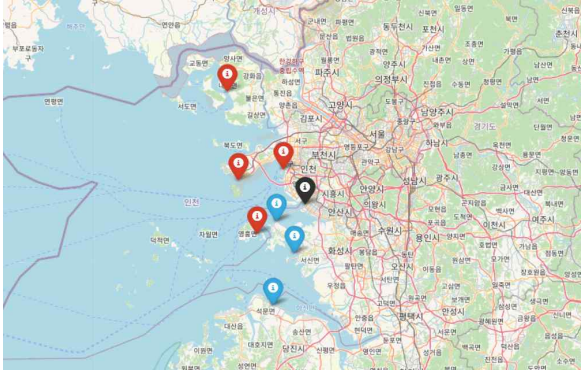


그림 9. (빨강) 인천지역 파출소, (파랑) 경기도 파출소, (검정) 신항만 파출소 (인천, 경기 교집합)

## 2. 파출소별 사고 확률 예측 시스템

본 제안서에서 실행한 앙상블 모델을 이용한 사고 유형 예측 시스템을 개선하여 파출소마다 자주 발생하는 사건 및 원인, 날씨, 계절, 위치 등 여러 요인을 고려해 사고 유형에 따른 확률을 예측한다. 이는 머신러닝 모델뿐만 아니라 딥러닝 모델을 이용해 입력값으로 고려 요인이 들어온다면 여러 은닉층을 거쳐 출력층에서 Softmax 함수를 통해 각 사고원인에 따른 확률을 도출이 가능할 것이다. Softmax 함수는 전 노드에서 온 값을 함수를 통해 확률값으로 변환한다. 이를 각 파출소에서 확인 후 모니터링이나 순찰 등 효율적인 업무를 도울 것이다. (그림 10) 예시로 **강릉 파출소는 부주의 사고다발구역을 평소보다 더 순찰한다. 이 예측 시스템은 단순 참고용일 뿐 평소 순찰 시스템을 기본으로 가져가되 확률이 높은 구역을 더 확인하는 것으로 한다.**

=소프트맥스 함수 예시=  
 1.0 ->softmax-> 0.0003  
 6.0 ->softmax-> 0.0417  
 4.0 ->softmax-> 0.0056  
 9.0 ->softmax-> 0.8381  
 2.0 ->softmax-> 0.0008  
 7.0 ->softmax-> 0.1134

표 3. 소프트맥스 함수 예시

강릉 파출소의 사고유형별 확률		
사고유형	확률	사고다발구역
부주의	60%	안록 xx 구역
조석 미인지	10%	안록 xx 구역
음주	20%	경포 xx 구역
자살	10%	경포 xx 구역

그림 10. 파출소별 사고 확률 예측 시스템 예시

## 3. 사고 원인별 대체 방안

사고 유형에는 추락익수, 고립, 기타익수, 추락, 표류, 수상레저 입수 등으로 (그림 12) 사고 원인으로서는 약 50% 부주의로 인한 사고들이다. 부주의 사고는 대부분 각종 안내 표지판의 준수사항을 지키지 않고 출입 통제 장소에 가서 사고를 당하는 경우가 많다. 준수사항을 잘 지킬 수 있도록 안전 수칙, 주의사항 같은 것들을 안내하는 표지판을 크게 세워 놓거나 방송으로 자주 안내한다. 본 제안서가 제안하는 사고원인별 대체 방안은 하나는 해수욕장 이용 시 코로나 사태 때 활용했던 전화 출입 인증과 같은 안전 수칙 준수에 대한 인증을 받고 출입하는 시스템으로 안전 수칙 인증 URL을 현수막이나 포스터, QR 코드로 제공하고 전화번호 인증을 통해 인증하는 것이다. 이후 안전 요원에게 인증을 확인하고 출입한다. 이는 개인에게 부주의에 대한 책임을 물으며 안전 수칙을 확인함으로써 부주의에 대한 사고를 줄일 수 있을 것이다.

또한 CCTV에 안전 수칙 준수 여부 인식 소프트웨어를 탑재해 미준수 시 안전 수칙 준수 안내 방송을 한다. 안전 수칙 준수 여부 인식 소프트웨어의 대표적 사례는 2021 수상작 중 딥러닝 기반 안전 장비 착용 탐지 솔루션과 같은 컴퓨터비전 알고리즘을 이용해 안전 수칙을 탐지하고 안내 방송을 한다.

조석 미인지로 인한 사고는 먼저 사람들이 많이 찾는 갯벌에 위 문단에서 소개한 안전 수칙 인증 시스템을 이용해 밀물 시간 때를 휴대전화로 알리를 보낸다. 만약 사고 발생 시 조석 미인지에 대한 책임을 물으며 이에 대한 처벌이나 관련 조치를 한다. 만약 인증 시스템을 사용하기 어려울 때 일정 수준 물이 차게 되면 입구 차단기를 내려 이동을 막는 시스템을 구축하여 사고를 예방한다. 예시로 최근 선재도에서 조석 미인지로 갯벌에 EV6, 랜드로버가 빠지는 사고가 일어났다. 선재도로 가는 길에 수위 감지 시스템을 설치하고 수위 변화에 따라 밀물, 썰물을 인식 후 선재도 입구, 출구의 차단기에 신호를 보내 작동한다. 밀물 때 사람, 차량이 이동 못 하도록 미리 차단해 사고를 예방할 수 있었을 것이다.

(그림 11) 하지만 기존 수위 감지 시스템의 고장률과 관리비가 많이 든다고 한다. 이는 딥러닝을 이용한 지능형 수위 감지 솔루션을 가진 기업이 존재하고 기존과 다른 수위 감지 시스템을 활용해 관

리 비용을 절감한다.



그림 11. 딥러닝 기반 수위 감지 시스템(아이브스)  
음주의 경우 안전 수칙 준수 인증 시스템의 경우 사용이 불가할 것이다. 따라서 사고다발구역에 펜스를 이용해 음주로 인한 추락을 방지한다. 자살의 경우는 한강 다리에 있는 것처럼 자살 예방 문구 또는 상담 센터와 바로 소통을 할 수 있는 스피커와 마이크를 설치한다. 자살 사고 의심의 경우 상담 센터에서 CCTV와 설치된 스피커, 마이크로 상담을 진행하고 최악의 상황을 대비해 파출소 출동의 시간을 벌고 상담 센터의 도움을 받도록 한다.

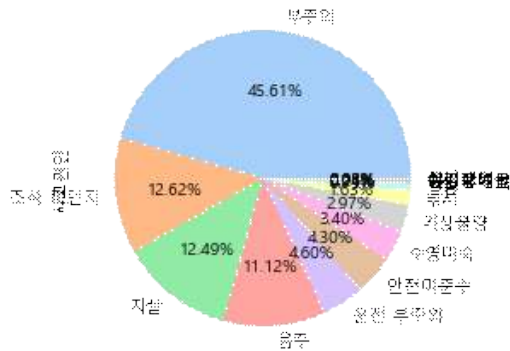


그림 12. 사고원인별 비율.

## V. 결론

본 제안서는 2017년~2020년도 연안 사고 데이터를 기반으로 파출소 업무분담 및 조정, 파출소별 사고 확률 예측 시스템, 사고원인별 대체 방안이라는 아이디어를 제안한다. 파출소 업무 분담 및 조정은 대표적으로 “신항만 파출소”나 “하늘바다 파출소”처럼 출동량이 적거나 많은 파출소 간의 업무 분담으로 효율적인 출동을 통해 연안 사고 대처에 가담한다. 파출소별 사고 확률 예측 시스템은 딥러닝 모델을 이용해 사고 유형별 사고 발생 확률을 알려준다. 각 파출소는 사고 발생 확률이 높은 유

형을 확인하고 유형별 자주 발생하는 장소를 주의 깊게 순찰하며 사고를 예방하도록 한다. 사고원인별 대체 방안으로는 유명 관광지나 해수욕장 이용 시 안전 수칙 준수 인증 시스템 활용을 한다. 개인이 수칙에 준수하겠다는 책임을 물어 처벌이나 관련 조치를 하고, 안전 수칙 준수 여부 인식 시스템으로 인식 후 미준수 시 방송을 한다. 조석 미인지는 안전 수칙 준수 인증 시스템을 유명 관광지에 사용되 사용 불가할 경우 딥러닝 기반 수위 감지 시스템으로 미리 출입을 차단하고 안내 문자나 방송으로 대피를 유도한다. 음주, 자살, 기타원인엔 펜스와 상담소와 즉석 상담 시스템을 활용한다.

충청남도, 인천광역시, 전라남도, 제주도 (생략) 울산 순으로 사고가 발생하는 것을 확인했다. 하늘바다 파출소는 다른 파출소에 비해 2배 정도 출동량이 많다. 같은 지역의 신항만 파출소는 그에 비해 현격히 적은 출동량을 가져가고 있다. 사고 유형은 추락익수, 고립, 기타익수, 추락 (생략) 순으로 사고가 일어나고 대부분 부주의, 조석 미인지, 자살, 음주에 대한 원인이다. 계절 기준으로 여름에 가장 많은 사고가 발생하며 가을, 봄, 겨울 순 해안을 기준으로 서해 남해, 동해 제주 순으로 사고가 자주 발생한다. 위와 같은 결과로 파출소 업무 분담 및 조정, 사고 원인별 대체 방안 아이디어를 도출하게 됐다.

직접 시도한 사고 유형 분류 모델링의 경우 시계열성을 모두 제거하여 정확하지 않다는 것을 염두에 뒀다. 단순 사고 유형 예측을 시행했을 때 Random Forest가 가장 성능이 좋았고 이는 파출소별 사고 확률 예측 시스템에 활용할 수 있다는 것을 보여준다.

데이터를 기반으로 연안 사고 예방 아이디어를 도출해봤다. 전체적으로 부주의에 대한 원인이 많다는 것에 해양경찰의 노력도 필요하지만, 시민들의 자발적 안전 확보도 필요하다고 생각한다. 이번 코로나 제한이 풀리고 많은 사람이 해수욕을 즐기러 관광을 떠날 것이다. 강릉이라는 바다 지역에 사는 학생으로서 내가 제안한 아이디어를 활용해 강릉뿐만 아닌 모든 바다 관광지가 안전하고 쾌적한 관광지가 되었으면 하는 바람이다.

감사합니다.