

119 신고 건수 예측을 위한 CatBoost 모델링과 활용방안 제안

팀 투빅스 : 송수현 · 조하늘 · 장가람 · 방지현 · 서동혁 · 강연주

Abstract

최근 기상 이변으로 인해 119 신고 건수가 급증함에 따라 소방 인프라의 부담이 가중되고 있다. 본 연구는 소방 데이터와 날씨 빅데이터를 융합하여 부산시의 119 신고 건수를 예측하는 모델을 개발하고, 이를 실무에 활용할 방안을 제시한다. 데이터 분석 과정에서 COVID-19, 지역별 지형 특성 등 외부 요인을 반영한 파생 변수를 생성하였고, 여러 트리 기반 모델의 성능을 비교 분석한 결과 CatBoost 모델이 가장 우수한 예측력을 보였다. 특히, 카운트 데이터의 특성을 고려하여 Poisson 손실 함수를 적용했을 때 모델의 성능이 더욱 향상되었다. SHAP 분석을 통해 '상세 주소(sub_address)'와 '자연재해 여부(disaster_day)'가 예측의 핵심 변수임을 확인하였으며, 이를 바탕으로 실시간 예측 및 시각화가 가능한 웹 대시보드를 개발했다. 최종적으로 본 연구는 예측 모델을 활용한 스마트 알람 시스템, 대화형 챗봇, 그리고 효율적인 소방 자원 배치 방안을 제안하여 도시 안전 관리 체계의 디지털 전환에 기여하고자 한다.

Key Words

119 신고 건수 예측, 기상 빅데이터, CatBoost-Poisson, 공간 특성, SHAP, 재난 대응 대시보드

1. 서론

최근 소방청 통계에 따르면 전국 119 신고 건수는 2017년 1,155만여 건에서 2022년 1,255만여 건으로 5년 새 100만 건 이상 늘어났다. 특히 2022년에는 전년 대비 약 47만 건(↑3.9%)이 증가하며 '하루 평균 3만 4천 건, 3초당 1건' 수준에 근접했다.¹⁾ 부산만 보더라도 2023년 신고가 75만 건을 넘어 '42초에 한 번' 꼴로 접수되는 등 대도시 소방 인프라가 이미 한계치에 도달하고 있다.²⁾

폭염 일수는 과거(1980-2000년) 대비 최근(2001-2023년) 평균적으로 증가하고, 발생 기간도 6-9월로 확대되는 추세이다.³⁾ 이처럼 기상 이변·기후 급변은 '전력·보건·안전' 전 분야에 부담을 주고 있으며, 소방서비스 역시 예외가 아니다.

따라서 소방데이터와 날씨 빅데이터를 융합한 119 신고 건수 예측 연구의 필요성이 증가하게 되었다.

소방데이터와 날씨 빅데이터를 융합한 119 신고 건수 예측을 주제로 실제로 활용할 수 있는 모델을 만들어 보았다. 흐름은 그림 1과 같다.

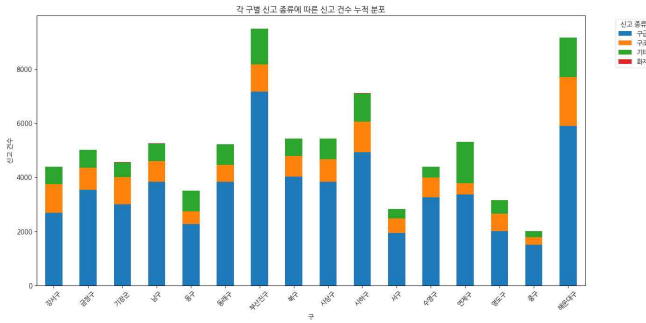
2.1 결측치 및 중복 데이터 확인

기상 데이터 분석 결과, 습도, 기온, 풍속, 강수량 변수에서 총 4,624건의 결측치가 확인되었다. 해당 변수들은 일반적으로 시간적으로 연속적인 특성을 가지며, 인접한 시점의 관측값과 높은 상관성을 보이는 경향이 있다. 특히, 전일의 기상 조건이 당일 기상에 일정 부분 영향을 미치며, 동일 지역 내 다른 시점의 관측값과도 유사한 패턴을 보일 수 있다는 점을 고려하였다. 이에 따라 본 연구에서는 ffill() 기법을 적용하여 전 행의 데이터로 결측값을 보완하였다.

2.2 EDA

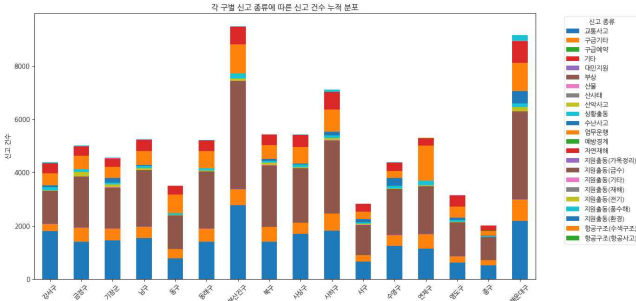


[그림 1] 분석 과정



[그림 2] 구별 신고 종류에 따른 신고 건수 누적 분포

1) 유성원, 2024, [2024년 소방청 통계연보](#), 소방청
2) 이현지, 2024, [\[보도자료\] 부산소방재난본부, 2023년 하루 42초마다 전화벨](#), 부산광역시 소방재난본부 종합상황실
3) 기상청, 2024, [\[보도자료\] 2024년 여름철 기후특성](#), 기상청 본청 기후변화감시과



[그림 3] 구별 신고 종류에 따른 신고 건수 누적 분포

그림 3은 구별 세부 신고 종류에 따른 신고 누적 분포를 나타냈다. 부상 23541건, 교통사고 18184건, 업무운행 7828건, 구급 기타 5904건으로 높은 신고 건수를 보였다.

2.3 추가 데이터 수집

추가 데이터는 행정동별 주민등록 인구 및 세대 현황과 Big-데이터웨이브에서 일별 행정동 시간 생활인구 월별 일 평균과 일별 행정동 연령 생활인구 월별 일 평균을 활용하였다. 두 데이터 모두 다음달 1일에 발표가 나는 데이터이기 때문에 DataLeakage에 위반되지 않게 전 월 데이터로 사용하였다. 인구 현황 데이터는 지역별 고령 인구수(60세 이상)가 영향을 미친다고 판단하여 feature에 포함하였다. 유동 인구 데이터는 유동 인구가 많은 지역에서 119 신고 건수와 연관이 있다고 판단하여 추가하였다.

추가변수	정의	출처	데이터 공표 일시
popu	일별 행정동 연령 생활인구 월별 일평균	부산광역시(민간 구매데이터)	다음 달 1일
population	행정동별 주민등록 인구 및 세대현황	행정안전부(주민 등록인구통계)	매월 1일 12시
old_population	행정동별 주민등록 인구 및 세대현황 중 60세 이상 인구수	행정안전부(주민 등록인구통계)	매월 1일 12시

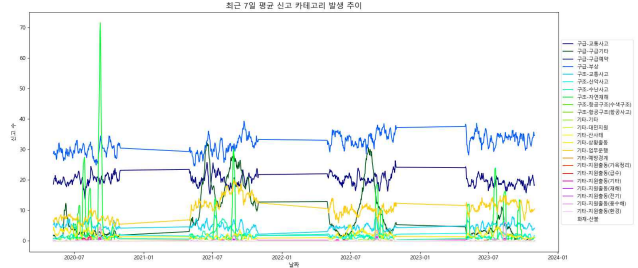
[표 1] 추가 입력 변수 데이터 설명

2.4 파생변수 생성

모델의 성능 향상과 해석력 강화를 위해, 본 연구에서는 도메인 지식에 기반한 파생변수를 설계해 적용했다. 주요 파생변수는 표2와 같고, 다음과 같은 판단 근거에 기반하여 도출되었다.

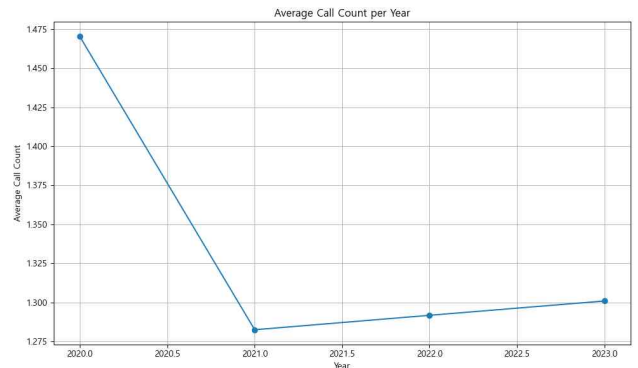
변수 종류	변수 이름	정의	출처
시간 파생변수	year, month, day, weekday	연, 월, 일, 요일	파생변수
기상 파생변수	wind_mean, hm_range, ta_max_3d_ma, hm_max_3d_ma, rn_day_bin	풍속 평균, 습도 범위, 최고기온 3일 이동평균, 최대습도 3일 이동평균, 강수량 구간화	파생변수
특수 상황변수	disaster_day, is_covid_year	자연재해 여부(이진), 코로나 영향권 여부(이진)	파생변수
지역그룹 변수	region_coastal, region_urban, region_mountain, region_lowland	해안, 도심, 산간, 평지 여부(이진)	파생변수

[표 2] 파생변수 데이터 설명



[그림 4] 최근 7일 평균 신고 카테고리 발생 추이

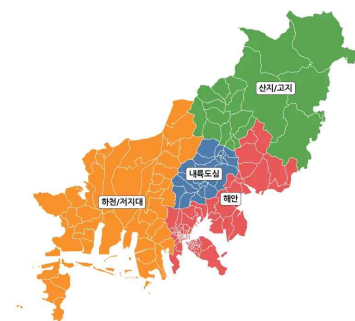
그림4는 신고 유형별 발생 추이를 시계열로 분석한 결과를 나타낸다. 분석 결과, 특정 시점에 구조-자연재해(연두색) 관련 신고 건수가 급격히 증가하는 이상 패턴이 반복적으로 관측되었다. 이는 특정 기상 조건에서 비정상적인 수요가 집중되는 이벤트성 현상의 가능성을 시사한다. 이에 따라, 해당 특성을 모델이 인식할 수 있도록 신고 유형 중 ‘자연재해’ 여부를 이진 변수로 재정의하여 포함하였다.



[그림 5] 연도별 평균 신고 건수

그림 5는 연도별 평균 신고 건수를 시각화한 결과이다. 분석 결과, 2020년도에 신고 빈도가 다른 연도에 비해 비정상적으로 높게 나타나는 경향이 확인되었다. 이는 COVID-19의 확산으로 인해 일반 병원의 응급실 기능이 일시적으로 제한되고, 이에 따라 구조요청이 119에 집중된 결과로 해석된다. 이러한 외부적 영향 요인을 모델이 학습할 수 있도록 ‘2020년 여부’를 구분하는 이진 파생변수를 도입하였다. 본 변수는 연도별 구조적 비정상성을 보완하고 모델의 일반화 성능을 높이는 데 기여한다.

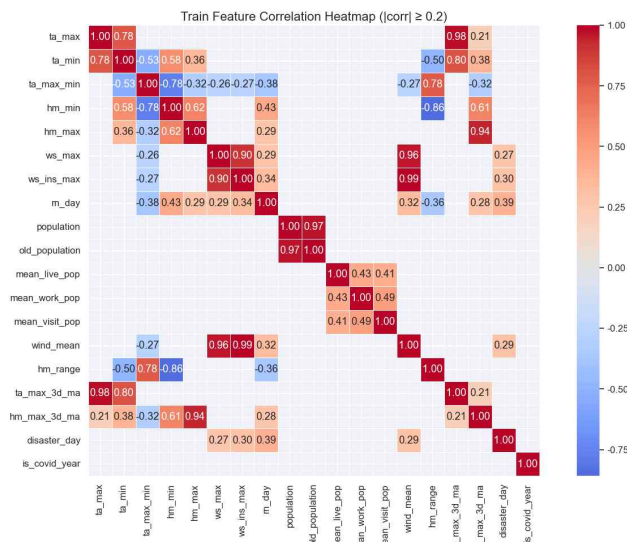
부산시 구 그룹별 지도



[그림 6] 부산시 구 그룹별 지도

부산 시는 지역별로 해안, 내륙 도심, 산지/고지대, 하천/저지대 등 지형적 특성이 뚜렷하며, 이는 강수량, 풍속 등 기상 지표에 대한 반응성과 신고 유형의 분포에 실질적인 영향을 미친다. 본 연구에서는 행정구(address_gu)를 기준으로 이들 지역을 4개의 그룹으로 분류하고, 해당 정보를 범주형 변수로 모델에 포함하였다(그림 6 참조). 이를 통해 동일한 기상 조건에서도 지역별로 다른 신고 반응을 모델이 인식할 수 있도록 하였으며, 공간적 이질성을 반영한 설명력 있는 예측 모델을 구축하는 데 목적이 있다.

2.5 상관 분석



[그림 7] 학습 변수간 상관관계 히트맵

그림 7는 기존 변수와 파생 변수를 포함한 데이터셋의 상관관계를 나타낸다. 분석의 용이성을 위해 상관관계의 절댓값이 0.2 이상인 경우만 표시하였다. 신고 건수(call_count)는 독립변수들과 전반적으로 낮은 상관계수를 보였으며, 평균 풍속(wind_mean)의 경우 0.169, 일 최대 순간 풍속(ws_ins_max)의 경우 0.166, 일 최대 풍속(ws_max)은 0.165를 기록하였다.

일 최대 순간 풍속(ws_ins_max)과 평균 풍속(wind_mean) 간에는 0.986의 매우 높은 상관관계가 나타났는데, 이는 풍속의 평균과 최대 풍속이 거의 동일한 패턴을 보이기 때문으로 판단된다. 일 최대 순간 풍속과 평균 풍속 간 관계를 회귀 분석한 결과, 일 최대 순간 풍속 F 통계량이 매우 큰 값을 기록하여서 일 최대 순간 풍속이 종속변수와 거의 완벽하게 일치한다고 판단하였다. 또한, 다른 변수 14개와 다중공선성 문제가 발생하였다. 문제 해결하기 위해 일 최대 순간 풍속을 변수에서 제거하였다.

또한, 일 최대 기온(ta_max)과 최고기온 3일 이동평균(ta_max_3d_ma) 간에 0.976의 매우 높은 상관관계가 확인되었다. 회귀 분석 결과 F 통계량이 매우 크게 나타났고, 14개 변수와 다중공선성 문제가 발생하여 최고기온 3일 이동평균을 변수에서 제외하였다. 마찬가지로 일 최대 습도(hm_max)와 최고습도 3일 이

동평균(hm_max_3d_ma)은 0.938의 높은 상관관계를 나타내어 최고습도 3일 이동평균을 제거하였다.

인구수(population)와 고령인구(old_population) 간에는 0.967의 높은 상관관계가 나타났으나, 전체 인구에서 고령인구를 제외한 인구와 고령인구는 밀접한 관련성을 가지면서도 해석상 구별되는 의미를 지니므로 두 변수 모두 유지하였다.

최종 address_gu, sub_address, stn, ta_max, ta_min, ta_max_min, hm_min, hm_max, ws_max, rn_day, population, old_population, 평균주거인구수, 평균직장인구수, 평균방문인구수, year, month, day, weekday, wind_mean, hm_range, disaster_day, is_covid_year, region_group 변수를 사용하였다.

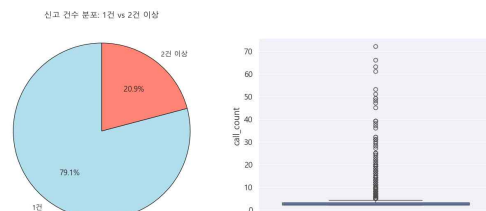
변수 종류	변수 이름	정의	출처
타겟 변수	call_count	119 신고 건수	기본 제공데이터
주소 데이터	address_gu, sub_address, stn	주소(구, 동), 관측소	기본 제공데이터
기상 데이터	ta_max, ta_min, ta_max_min, hm_min, hm_max, ws_max, m_day	최고 기온, 최저기온, 일교차, 최저습도, 최대습도, 최대풍속, 최대풍속, 일강수량	기본제공 데이터
시간 파생변수	year, month, day, weekday	연, 월, 일, 요일	파생변수
기상 파생변수	wind_mean, hm_range	풍속 평균, 습도 범위	파생변수
특수 상황변수	disaster_day, is_covid_year	자연재해 여부(이진), 코로나영향권여부(이진)	파생변수
지역그룹변수	region_coastal, region_urban, region_mountain, region_lowland	해안, 도심, 산간, 평지 여부(이진 변수)	파생변수
인구변수	population, old_population, mean_live_pop, mean_work_pop, mean_visit_pop	전체 인구, 60세이상 인구, 평균주거인구수, 평균직장인구수, 평균방문인구수	행정안전부, Big-데이터웨이브(부산시)

[표 3] 전체 입력 데이터 변수 설명

2.6 119 신고 건수 예측 알고리즘 개요

a. 모델링 방법론 소개

1) 클래스 불균형 문제



[그림 8] 신고 건수 1건과 2건이상 비교

전체 61,771건의 신고 데이터 중 신고 건수 1건이 79.1%를 차지하며, 이는 심각한 클래스 불균형(class imbalance) 문제를 야기한다. 이로 인해 모델은 주로 다수 클래스(1건)에 과적합되며, 신고가 많은 날에 대한 예측력이 저하되는 경향을 보인다. 이러한 편향을 완화하기 위해 CatBoost와 같은 Gradient Boosting 계열 모델에서는 가중치 기반 손실 함수(Weighted Loss Function)를 적용할 수 있다. 이 방식은 희귀한 고신고일에 대한 오차를 더 크

게 반영함으로써 소수 클래스 학습을 강화하고, 불균형한 분포에서도 예측 성능의 균형을 확보하는 데 효과적이다.

2) 지역 기반 예측 모델

119 신고 건수 예측의 정확도를 높이기 위해 지역별 기상 조건과 지형적 특성의 차이를 반영한 공간 분할 전략을 도입하였다. 지역을 단일한 집합으로 간주할 경우, 기상 변수에 대한 민감도 차이가 모델링 과정에서 희석될 수 있으며, 이로 인해 지역 특화 패턴을 포착하는 데 어려움이 발생한다. 따라서 지역의 기후적, 지형적, 도시 구조적 특성을 기준으로 다음 네 가지로 구분하였다.

지역 그룹	포함 구	특성 요약
region_coastal	해운대, 수영, 영도, 중구, 동구, 서구, 남구	해풍, 습도, 태풍 영향
region_urban	부산진, 동래, 연제	도시 열섬, 고온
region_mountain	금정, 기장	강수 편차, 기온 낮음
region_lowland	강서, 북구, 사상, 사하	안개, 침수, 습도

[표 4] 지역별 변수 특성

이러한 지역 구분은 단순 행정구 단위의 경계보다 실제로 기상 요인과 신고 원인 간의 인과관계를 더 잘 설명해줄 수 있다.

지역을 해안, 도심, 고지대, 저지대로 구분하여 예측 모델을 개별적으로 학습한 결과, 각 지역의 지형적 특성과 환경 요인에 따라 중요하게 작용하는 변수에 뚜렷한 차이가 나타났다. 변수 중요도 분석은 SHAP 값을 기반으로 수행되었으며, 지역별 상위 4개 변수는 다음과 같다.

	1순위	2순위	3순위	4순위
region_coastal	ws_max	ws_ins_max	address_gu	cat
region_urban	ws_max	cat	address_gu	year
region_mountain	sub_address	day	year	ta_max
region_lowland	ws_ins_max	sub_address	cat	address_gu

[표 5] 지역별 변수 중요도

b. 트리 기반 모델

기상 및 지역 정보를 활용한 119 신고 건수 예측 문제를 해결하기 위해 트리 기반 앙상블 모델을 주로 활용하였다. 이는 기상 변수와 신고 건수 간의 관계가 단순한 선형성이 아닌 복잡한 비선형적 상호작용을 포함하고 있으며, 특정 조합의 변수 조합(예: 고온 + 강풍 + 주말)에 따라 신고가 급증하는 등 비정형적 패턴이 존재하기 때문이다. 트리 기반 모델은 이러한 특성을 효과적으로 포착할 수 있는 구조적 장점을 갖는다.

세 가지 대표적인 트리 기반 앙상블 모델인 XGBoost, LightGBM, CatBoost를 적용하여 성능을 비교하였다. 예측 성능

평가지표로는 평균 절대 오차(MAE), 평균 제곱근 오차(RMSE), 그리고 결정계수(R^2)를 사용하였다. 그 결과는 다음과 같다.

	MAE	RMSE	R2
XGBoost	0.46	1.28	0.05
lightGBM	0.44	1.03	0.38
CatBoost	0.43	0.99	0.42

[표 6] 트리기반 모델 비교

XGBoost는 MAE 0.46, RMSE 1.28, R^2 0.05로 예측 정확도와 설명력 모두에서 낮은 성능을 보였으며, 이는 해당 모델이 비선형적 패턴이나 지역별 기상 특성의 상호작용을 충분히 반영하지 못했음을 시사한다. 반면, LightGBM은 MAE 0.44, RMSE 1.03, R^2 0.38로 XGBoost보다 개선된 성능을 나타냈으며, Leaf-wise 성장 방식이 변수 간 상호작용 포착에 효과적으로 작용한 것으로 보인다. 가장 우수한 성능을 보인 CatBoost는 MAE 0.43, RMSE 0.99, R^2 0.42로 신고건수의 실제 분포를 가장 잘 반영하였다. 특히 범주형 변수 처리와 순열 기반 학습 방식이 예측 성능 향상에 기여하였으며, 결측치와 이상값에 대한 강인성 측면에서도 본 데이터에 적합한 모델로 평가된다.

종합하면, 세 모델 중 CatBoost가 가장 안정적이고 설명력 있는 예측 성능을 보여주었으며, 이는 향후 실제 정책적 의사결정이나 재난 대응 시스템에 적용 가능한 모델로서의 가능성을 높여준다.

c. 최종 모델 선정

최종 모델은 CatBoost에 Poisson 손실 함수를 적용하여 학습되었다. 이는 신고 건수가 0 이상의 이산형 카운트 데이터라는 특성을 반영한 것으로, RMSE 기반 손실 함수보다 더 적합하고 안정적인 예측을 가능하게 한다. 성능 비교 결과, 두 모델 모두 MAE는 0.43으로 동일하였으나, RMSE는 Poisson 모델이 0.97로 RMSE 기반 모델(0.99)보다 낮았고, R^2 역시 0.45로 더 높았다. 이는 Poisson 손실 함수가 극단적인 오차를 효과적으로 제어하면서 모델의 설명력을 향상시켰음을 의미한다. 결과적으로, 트리 기반 모델의 구조적 장점에 Poisson 손실의 분포 적합성을 결합한 접근은 카운트형 예측 문제에 있어 효과적인 전략으로 판단된다.

	MAE	RMSE	R2
CatBoost - RMSE	0.43	0.99	0.42
CatBoost - Poisson	0.43	0.97	0.45

[표 7] CatBoost 모델 비교

2.7 119 신고 건수 예측 알고리즘 모델링

CatBoost 모델의 성능을 추가적으로 향상시키기 위해 Optuna 프레임워크를 활용하여 주요 Hyper Parameter에 대한 베이지안 최적화 기반 튜닝을 수행하였다. Optuna는 자동화된 탐색 알고리즘을 통해 탐색 공간 내에서 성능이 가장 우수한 조합을 효과적으로 탐색할 수 있는 고성능 하이퍼파라미터 최적화 도구이다.

최적화 대상은 모델의 일반화 성능에 크게 영향을 미치는 주요 파라미터들이며, 그 결과 최종 선정된 최적 하이퍼파라미터 조합은 다음과 같다.

Parameter	Value
loss_function	Poisson
iterations	1172
depth	6
learning_rate	0.0038166161833928687
l2_leaf_reg	6.178500608595952
bagging_temperature	0.4169323820127707
random_strength	2.6137903340558464
border_count	211

[표 8] 최적의 파라미터

이러한 파라미터 설정을 통해 CatBoost 모델은 보다 안정적이고 설명력 있는 예측 성능을 확보하였다. 결과적으로 Optuna 기반의 자동화된 튜닝은 모델 성능 향상뿐 아니라, 수작업 기반의 반복적 실험 과정을 효율화하는 데에도 기여하였다.

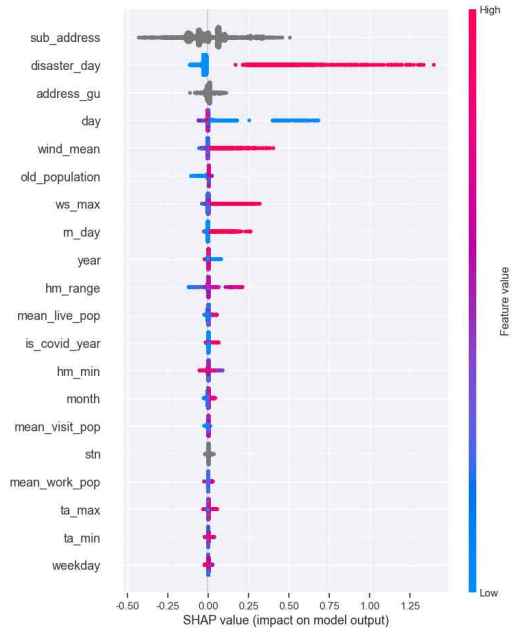
3. 결론

3.1 결론

모델의 예측 성능을 확인한 이후, 모델의 해석 가능성을 확보하고, 각 입력 변수들이 예측 결과에 어떤 방식으로 영향을 미쳤는지를 정량적으로 분석하기 위해 SHAP(Shapley Additive Explanations) 분석을 수행하였다.

최종 선정된 CatBoost - Poisson 모델에 대해 SHAP 분석을 적용함으로써, 신고 건수 예측에 가장 큰 영향을 미치는 주요 변수들을 식별하고, 지역별 기상 특성과 신고 유형 간의 상관성을 해석적으로 제시하고자 한다. 이는 단순한 예측 정확도 이상의 정책적 해석력과 실무 적용 가능성을 제공하는 데 목적이 있다.

CatBoost-Poisson 모델에 대해 SHAP을 활용한 변수 중요도 분석을 수행한 결과, sub_address와 disaster_day가 예측에 가장 큰 영향을 미치는 변수로 나타났다. sub_address는 지역별 신고 패턴 차이를 반영하며, disaster_day는 자연재해 발생 시 신고 건수가 증가하는 경향을 명확히 보여주었다. 이외에도 address_gu, day, wind_mean, ws_max, rn_day 등의 변수들이 중간 수준 이상의 기여를 하며, 강풍과 강수와 같은 기상 조건이 신고 발생의 주요 요인임을 시사하였다. 반면, ta_max, ta_min, month, weekday 등의 기온·시간 변수는 상대적으로 영향력이 낮았고, 인구 관련 변수들도 제한적인 예측 기여도를 보였다. 이는 기상 재해와 지역성이 신고건수 예측에 핵심적인 역할을 한다는 점을 정량적으로 보여준다.



[그림 9] SHAP Value 활용 변수 중요도 분석

3.2 활용방안

본 연구에서는 부산시의 기상 데이터와 119 신고건수를 통합적으로 분석하고 시각화 및 예측할 수 있는 웹 기반 대시보드 시스템을 설계·구현하였다.

Streamlit을 활용하여 개발된 본 시스템은 기능별로 두 개의 모듈로 구성되며, 각각 지역 단위의 상세 조회 및 예측 기능, 부산시 전체의 신고 및 기상 현황에 대한 종합 시각화 기능을 제공한다.

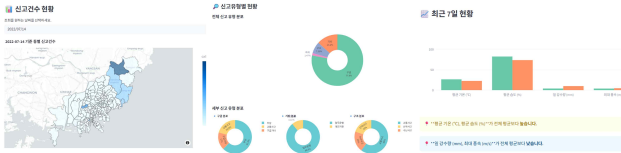


[그림 10] Streamlit을 활용한 웹 기반 대시보드 시스템

첫 번째 모듈(구 단위 신고건수 조회 및 예측 도구)은 사용자가 특정 날짜와 행정구를 선택하면 해당 지역의 주요 기상 지표(기온, 습도, 강수량, 풍속 등) 및 신고 건수를 요약한 메트릭 정보를 제공한다. 아울러, 동별 신고 건수 분포를 막대 그래프로 시각화하여 지역 간 편차를 직관적으로 파악할 수 있도록 하였다. 또한, 최근 1주일 또는 1개월간의 평균 기온, 강수량, 습도, 풍속 변화 추이를 시계열 그래프로 구현함으로써, 기상 패턴의 시간적 흐름 분석을 가능하게 하였다. 나아가, 사용자 입력 기반의 기상 변수에 대해 기계학습 모델을 통해 신고 건수를 예측할 수 있는 기능도 탑재되어 있어, 지역 행정기관의 사전 대응 계획 수립에 유용하게 활용될 수 있다.

4) 첫 번째 모듈(구 단위 신고건수 조회 및 예측 도구)

<https://call-prediction-a9b2zjif8rd2agrvbvgygx.streamlit.app>



[그림 11] Streamlit을 활용한 웹 기반 대시보드 시스템

두 번째 모듈⁵⁾은 부산시 전체를 대상으로 하는 시각화 중심의 대시보드로서, 선택한 날짜 기준 동별 신고 건수의 공간 분포를 Choropleth 지도 형식으로 제공한다. 또한, 전체 신고 유형 및 세부 유형의 비율을 파이 차트로 시각화하고, 최근 7일간의 기상 지표 평균값을 전체 평균과 비교 분석함으로써 단기적 이상 기상 조건에 대한 인지 및 조기 경고의 기초 자료로 활용할 수 있다.

이와 같은 기능 구현을 통해 단순한 데이터 확인을 넘어, 시간적·공간적 관점에서 신고 패턴과 기상 요인 간의 상관관계를 시각적으로 탐색할 수 있으며, 예측 기반의 판단 자료로서 정책 기획과 실무 대응 모두에 실질적인 근거를 제공한다. 향후에는 실시간 API 연동을 통해 실시간 데이터 반영이 가능하며, 교통, 인구 밀도 등의 외부 데이터를 통합함으로써 정보 시스템, 자동화된 의사결정 지원, 모바일 연계 서비스 등으로의 확장이 가능할 것으로 기대된다. 본 시스템은 도시 안전 관리와 재난 예방 대응체계의 디지털 전환을 위한 기초 플랫폼으로서 실용적 가치를 지닌다.

해당 시스템은 2020년부터 2023년까지의 5~10월 기간동안 수집된 학습용 데이터를 기반으로 구현되었으며, 현재는 기본 CatBoost 모델을 활용하고 있다. 향후에는 모델 고도화 및 실시간 예측 성능 개선을 통해 정확도를 향상시킬 수 있을 것으로 기대된다.

※ 본 시스템의 웹 구현 예시는 아래 링크를 통해 확인 가능하다.

3.3 아이디어 제안

미국의 국가재난대응체계(NIMS), 국내 재난관리 매뉴얼, 그리고 각종 재난 대응 연구⁶⁾에서는 신고 대응의 핵심 요소로 다음 세 가지를 공통적으로 강조하고 있다. 소통 및 정보 관리, 자원 관리, 지휘 및 조정이다. 이에 부합하는 세 가지 아이디어를 다음과 같이 제안한다.

첫째, 119 신고 건수 예측과 기상 정보를 활용한 스마트 알람 시스템 도입이다. 현재 부산시는 119 출동 정보와 기상특보를 각각 제공하고 있으나, 119 신고 건수 예측을 기반으로 한 알람 시스템은 도입되어 있지 않다. 본 시스템은 기상 정보와 119 신고 건수를 예측하는 모델을 기반으로, 일정 단계 이상의 신고 건수가

예측될 경우 사용자 위치를 반영해 스마트 알람을 발송한다. 사용자는 해당 알람을 통해 예상되는 안전사고를 사전에 인지하고, 예방 조치를 취함으로써 실제 신고 건수를 줄이는 효과를 기대할 수 있다. 기상 정보와 예측 결과를 융합하여, 특정 지역에서 신고 건수 급증이 예상되는 경우 실시간으로 알람을 제공하는 시스템을 구축하고자 한다. 2022년 서울시에 시범 도입된 ‘서울시 스마트 재난관리 시스템’은 이러한 선진적 접근의 대표 사례로, 해당 시스템 도입 이후 서울시는 재난 대응 시간을 단축하여 화재 및 기타 재난 발생 시 골든타임을 확보하였다.

둘째, 정보화를 통한 부산 시민과의 접근성 향상을 위해 LLM 기반 대화형 챗봇 개발을 제안한다. 이는 최근 국내 사용자 1,000만 명을 돌파한 챗GPT와 같이 높은 접근성과 활용도를 기대할 수 있다. 대화형 챗봇은 일정 기간의 신고 예측 건수 데이터를 바탕으로, 시민에게 방문 예정 지역의 위험도를 안내하여 사전 사고 예방 효과를 높일 수 있다. 신고 건수가 낮게 예측되어 스마트 알람이 발송되지 않은 경우에도, 시민이 궁금한 상황에 대해 챗봇을 통해 언제든지 정보를 얻을 수 있다. 또한, 소방청에서는 챗봇 도입을 통해 선제적 대응, 소방대원 인력 배치 등 업무 효율화 효과도 기대할 수 있다.

셋째, 위험 예상 지역에 적절한 인력 배치를 통한 자원 운영 최적화를 제안한다. 2025년 3월 산불 피해 당시, 해당 위험 지역에 소방 인력(헬기 128대, 군·소방 인력 7,000여 명)이 부족하다는 지적이 있었다⁷⁾. 이와 같이, 부산시 내 119 신고 건수 예측 결과를 기반으로 위험 기간 및 지역에 더 많은 인력과 소방차를 배치하는 등 자원 운영을 최적화하면, 위험 상황 발생 시 신속한 대응으로 도시 안전과 재난 대응에 긍정적인 영향을 기대할 수 있다. 인력 및 장비 배치뿐만 아니라, 각 지역 소방서의 근무 일정 조율을 통해서도 효율적인 결과를 얻을 수 있다. Streamlit 대시보드에서 높은 신고 건수가 예상되는 날에는 최소 인원보다 많은 인력을 배치하고, 신고 건수가 낮은 날에는 최소한의 자원만을 활용함으로써 소방대원의 업무 부담을 조절할 수 있다. 소방청 공식 자료⁸⁾에 따르면, 최근 5년간(2017~2021) 극단적 선택을 한 소방공무원은 67명(연평균 13.4명)이며, 이 중 상당수가 파로, 만성 피로, 정신적 스트레스를 원인으로 직무상 과중한 업무와 연관된 것으로 분석된다. 따라서, 예측 기반 자원 운영을 통해 인력과 장비를 보다 합리적으로 배분함으로써, 위험 상황 대응력을 높이는 동시에 소방공무원의 근무 환경 개선과 정신건강 보호라는 이중의 효과를 기대할 수 있다.

5) 두번째 모듈 (부산시 전체 종합 대시보드)

<https://call-monitoring-jincjgva3f5qh4n5gfcfi.streamlit.app>

6) 오영석, 2019, 미국의 국가재난대응체계 및 총괄조정방식 연구, 행정안전부

7) 김소진, 2025, 산불 점진 크게·수시로 나는데 인력 그대로...개인장비로 진화 웬말, 농민신문

8) 문현주, 2022, (보도자료) 소방청, 소방공무원 정신건강 직접 챙긴다, 소방청