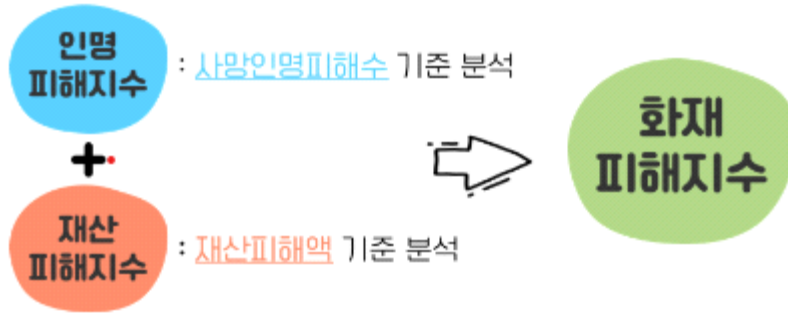


1) 아이템명(제목)

화재! 선제적 대응하자!

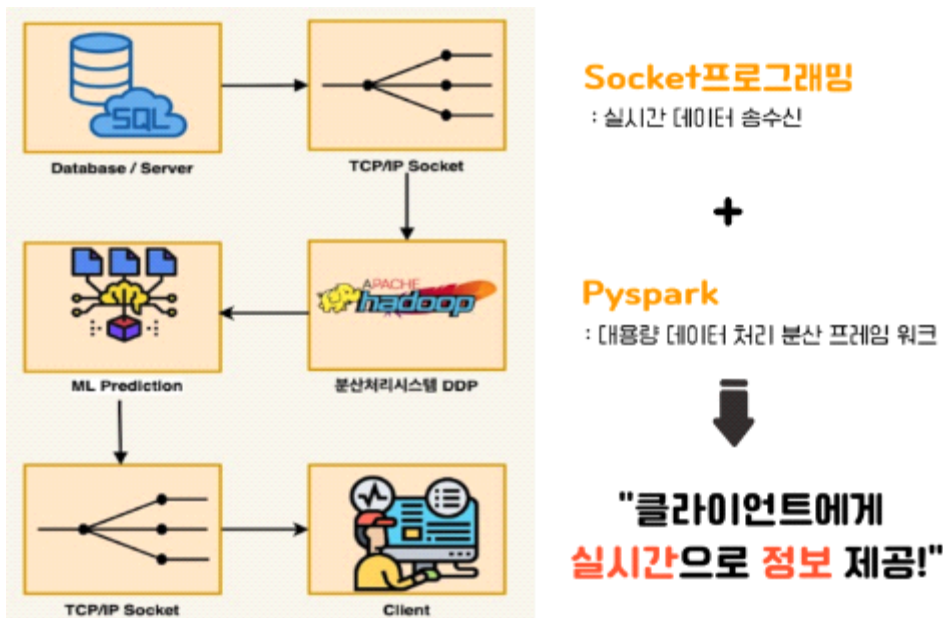
2) 활용 및 아이디어 요약

1. 화재 피해지수 도출



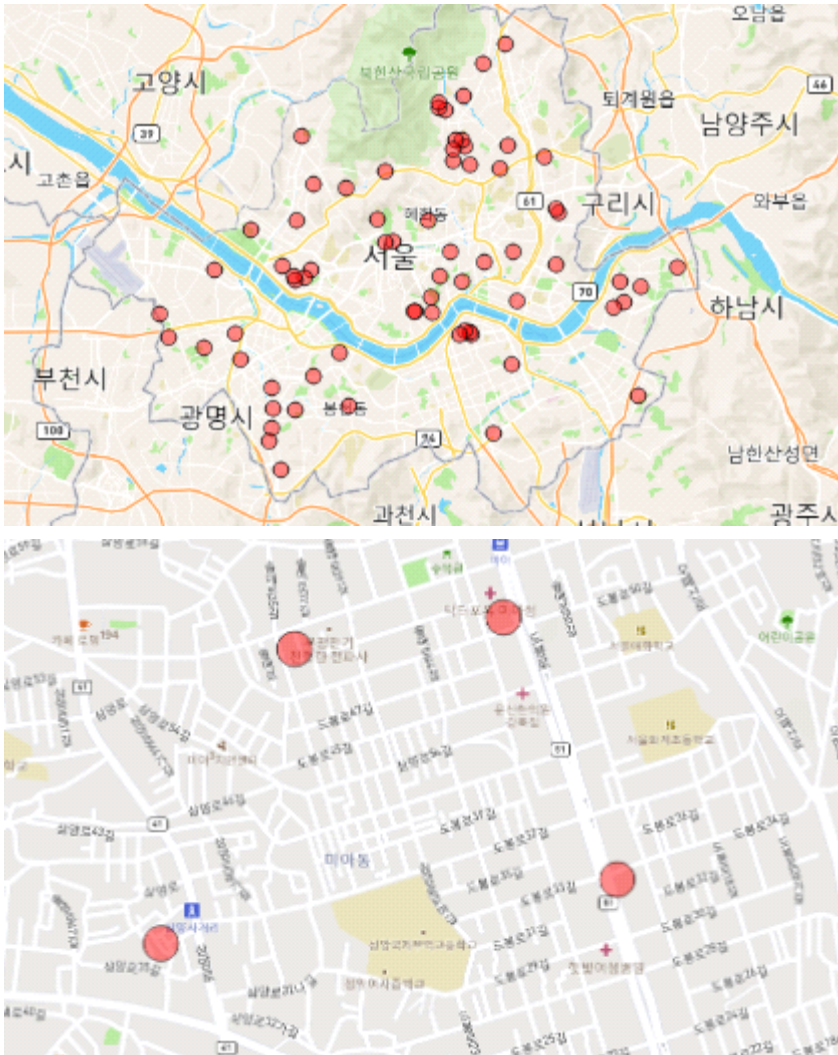
- "인명 피해 발생 확률"을 예측하는 모델
- "재산 피해액"을 예측하는 모델
- 두 모델의 예측 값을 합해서, 종합적인 화재 피해지수를 산출

2. 실시간 처리 시스템 구현



- 실효습도, 에너지 사용량 등 동적으로 활용할 수 있는 데이터들을 실시간 처리 시스템 구현

3. 태블로를 활용한 위험 건물 시각화



- 분석결과를 토대로 위험건물을 Tableau를 이용하여 시각화하였다.

3) 필요성 및 목적

우리나라의 지난 10년간(2012년~2021년) 연평균 화재 건수는 41,571건, 인명피해는 2,242명(사망 310명, 부상 1,933명), 재산 피해는 5,607억 원이다. 대한민국 공식 전자정부 누리집에 따르면 지난 10년간 화재에 의한 인명피해는 평균적으로 매년 2,300여 명이 발생하고 있다. 또한 재산피해액은 2012년 2,894억 원에서 2017년 5,069억, 2021년 1조 991억 원으로 해를 거듭할수록 큰 폭으로 증가하고 있음을 알 수 있다. 2021년도에는 건축/구조물 화재의 발생 건수가 23,997건으로 전체 화재 건수의 66.2%를 차지하고 있으며, 사망자 240명(87.0%), 부상자 1,589명(85.7%), 1,046,802백만 원(95.2%)의 재산 피해를 발생시킴으로써 유형별 화재 중 건축/구조물 화재가 매년 가장 큰 인명, 재산피해 규모를 차지하고 있다. 따라서 건물의 화재 위험도를 예측 및 활용할 수 있다면, 화재로 인한 물적/인적 피해를 줄이고 예방하는 데 큰 도움이 될 것이다. 이를 위해 '건물 안전성 데이터 마이닝 및 분석을 통해 건물의 위험성 지표 및 응용 가능한 프레임워크를 개발'했다. 이러한 지표를 건물 소유자, 시설 관리자, 정부 기관 등이 이용한다면 더욱더 안전한 사회를 만드는 데 기여할 수 있을 것으로 기대된다.

화재에 의한 피해는 크게 재산 피해와 인명피해로 구분된다. 화재 발생 시 발생하는 재산피해액(재산 피해)과 인명피해 발생여부(인명피해)를 예측하여 화재 위험성 지표(이하 화재피해등급)를 개발했다. 기존의 국내 연구를 살펴보면, 화재와 관련된 데이터 기반의 분석은 시도-시군구별로 화재위험지수를 산정해 오고 있다. 하지만 이러한 분석 결과는 지역 단위로 제공되기 때문에 실제 정책이나 화재안전점검 우선순위 선정 등 건물 단위의 예방 활동에 활용하기에는 어려움을 클 것으로 예상된다. 국외 사례를 살펴보면, 미국 아틀란타 소방청 (Atlanta Fire Rescue Department, AFRD)에서 구축한 Firebird가 대표적이다. Firebird는 건물 단위로 화재위험도를 예측하여 화재 발생 위험도가 높은 건물의 점검 우선순위를 결정하는 것이 주목적이다. 해당 모델은 랜덤포레스트(Random Forest) 기반으로 개발되어 약 71%의 정확도를 보였으나, 건물에 관한 속성도 면적, 층수, 필지 등 한정적인 변수만을 활용했다는 한계점이 존재한다. 따라서 본 연구에서는 소방재난본부, 국가공간정보포털 등에서 관리되고 있는 건물 데이터를 중심으로 다양한 변수들을 고려한 화재 관련 융합 데이터 셋을 생성하고 이를 통해 화재피해등급을 산정하는 연구를 진행했다. 또한 실시간으로 동적인 데이터를 수집하고 처리하기 위해 소켓 프로그래밍 및 분산 처리 시스템을 통하여 선제적인 대응이 가능한 시스템을 구축했다.

화재 발생 시 재산 피해액 및 인명피해 예측을 위해 사용한 변수들로 현장안전센터거리, 기온, 풍속, 가시거리, 실효습도, 전기사용량, 가스사용량, 화재발생시각 등이 고려되었다. 위에 나타난 대부분 변수는 화재가 발생하기 이전에 실시간으로 수집하고 업데이트할 수 있다. 즉, 화재피해등급 변화의 실시간 모니터링이 가능하며 대용량 데이터 처리에 견딜 수 있는 시스템을 구축했다. 이를 통해 건물 화재 발생에 대한 사전 대비가 가능하며 화재피해등급에 따라 화재의 피해 규모를 사전에 예측하고 인지할 수 있다. 화재 예방과 더불어 화재 발생 시 건물 소유자, 시설 관리자, 정부 기관 등이 신속하고 선제적인 대처를 하는 데 도움을 줄 수 있다.

4) 아이디어의 세부내용

건물 단위로 데이터를 융합하여 화재피해등급을 예측하는 것이 목적이기 때문에 전기/가스 사용량, 날씨 데이터, 화재 출동 현황 등 다양한 데이터를 건물별로 수집했다.

데이터명	설명	출처
전기에너지	2022년 건물별 전기에너지 사용량	국토안전부
가스에너지	2022년 건물별 가스에너지 사용량	
화재출동현황	사망인명피해수, 화재발생 시각, 현장안전센터거리, 시간단위날씨 (기온, 풍속), 가시거리, 재산피해액	서울 소방재난본부
상대습도	2017~2022년 시간별 상대 습도 데이터	기상청

데이터 전처리

- 변수 선택 : 결측치가 20% 이상인 칼럼을 제거하고 분석 목적에 맞지 않는 데이터를 분류했다.

- 실효습도 변수 생성 : 실효습도란 화재 예방을 목적으로 사용하기 위해 고안된 것으로 건조도를 나타낸다. 평균습도에 지나온 시간에 따른 가중치를 두어 산출하게 되고 건조한 날이 연속되는 경우 실효습도는 낮아지고 불이 날 가능성은 커지게 된다. 기상청에서 얻은 상대습도 데이터를 기준으로 계산하여 2017년~2022년의 실효습도 변수를 생성했다.

$$He = (1 - r) \times [H_t^{0d} + r H_t^{1d} + r^2 H_t^{2d} + r^3 H_t^{3d} + r^4 H_t^{4d}]$$

$$r = 0.7$$

$$* \text{실효습도}(He) = (1-r) \times [H0dt + r H1dt + r2 H2dt + r3 H3dt + r4 H4dt]$$

$$r = 0.7$$

He : 실효습도

H0d : 당일의 상대습도

H1d : 1일 전의 상대습도

H2d : 2일 전의 상대습도

H3d : 3일 전의 상대습도

H4d : 4일 전의 상대습도

결측치

- 5개의 변수(현장소방지역대거리, 시간단위기온, 시간단위풍속, 시간단위풍향, 시간단위습도)에서 각각 7, 3, 34, 34, 2개의 결측치를 발견했다.
- 현장소방지역대거리 : 위도, 경도를 통해 하버사인 거리를 계산하여 인접한 건물 값으로 대체했다.
- 기온, 습도 : 일정한 간격으로 타임스탬프가 찍힌 시계열 데이터가 아니기 때문에 보간법으로 대치가 불가하여 삭제했다.
- 풍속, 풍향 : 풍속과 풍향은 시간 별로 변화하는 폭이 크기 때문에 일별 평균 풍속과 풍향으로 대체하지 않고 삭제했다.

하버사인 거리

- 두 위경도 좌표 사이의 거리를 구할 때 사용
- 곡률의 영향을 최소화

$$d = 2r \arcsin \left(\sqrt{\text{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1) \cos(\varphi_2) \text{hav}(\lambda_2 - \lambda_1)} \right)$$

$$= 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right)$$

이상치

- 변수별 기술 통계량을 확인해 보고 이상치로 의심되는 값들이 존재하는지 확인했다.
- 총 6개의 변수(현장소방서거리, 현장안전센터거리, 시간단위가시거리, 재산피해액, 전기사용량, 가스사용량)에서 이상치로 의심되는 값들을 확인했다.
- 분포 형태에 따라 이상치 확인에 유리한 방법들이 다르다. 변수들의 분포가 다양했기에 여러 가지 방법을 활용하여 이상치를 탐지했다.
 - Graphical method : Box-plot, Scatter plot을 통해 시각적으로 이상치를 확인했다.
 - z-score : 각 변수를 표준화하여 z-score 값을 통해 3표준편차 이상의 값들을 확인했다.
 - 추가적으로 변수별 특징을 고려하여 결측치 대체 및 제거했다.
- 현장소방서거리, 현장안전센터거리, 시간단위가시거리
 - Box-plot과 z-score를 확인하여 각각 17, 10, 3000 이상의 값들을 이상치로 판별하고 거리를 계산하여 인접한 건물의 값으로 대체
- 재산피해액
 - 변수 특성상 제거하지 않는 것이 맞지만 분석의 정확도를 위해 Box-plot을 통해 확인한 6개의 이상치를 제거했다.
- 전기사용량(KWh), 가스사용량(KWh)
 - Box-plot과 z-score를 확인하여 각각 349개, 670개의 값들을 이상치로 판별하고 제거했다.

데이터 병합

- 화재출동현황 데이터셋을 기준으로 전처리한 데이터들을 병합했다.
- 전기에너지와 가스에너지는 주소를 기준으로 병합했고 실효습도 변수는 화재발생일자를 기준으로 병합했다.

	현장안전센터거리	시간단위기온	시간단위풍속	시간단위가시거리	실효습도	가스사용량(KWh)	전기사용량(KWh)	화재발생시각_새벽	화재발생시각_저녁	화재발생시각_주간
0	1.0	6.3	2.3	294	57.37625	3519.916667	3519.916667	0	0	1
1	2.0	7.6	2.1	409	57.37625	41868.750000	41868.750000	0	0	1
2	4.0	9.2	3.1	331	57.37625	8635.750000	8635.750000	0	0	1
3	2.0	8.3	3.6	305	57.37625	457670.916667	457670.916667	0	0	1
4	4.0	6.4	2.1	221	57.37625	4711.583333	4711.583333	0	1	0

모델링 기법

- 인명피해발생여부 - 로지스틱 회귀분석
 - 총 8개의 변수(현장안전센터거리, 시간단위 (기온, 풍속, 가시거리), 실효습도, 전기사용량(KWh), 가스사용량(KWh), 화재발생시각)를 바탕으로 인명피해발생여부에 대해 모델링을 진행했다.
 - 각 변수별 결과는 다음과 같았으며, **예측 정확도는 93.9%**를 기록했다.

	coef	std err	z	P> z	[0.025	0.975]
현장안전센터거리	-0.1501	0.035	-4.323	0.000	-0.218	-0.082
시간단위기온	0.1644	0.056	2.945	0.003	0.055	0.274
시간단위풍속	0.0921	0.041	2.221	0.026	0.011	0.173
시간단위가시거리	0.5530	0.049	11.348	0.000	0.458	0.649
실효습도	-0.1893	0.047	-4.070	0.000	-0.280	-0.098
가스사용량(KWh)	-0.0263	7.69e+05	-3.43e-08	1.000	-1.51e+06	1.51e+06
전기사용량(KWh)	-0.0263	7.69e+05	-3.43e-08	1.000	-1.51e+06	1.51e+06
화재발생시각_새벽	-2.3421	0.076	-30.996	0.000	-2.490	-2.194
화재발생시각_저녁	-3.0584	0.086	-35.625	0.000	-3.227	-2.890
화재발생시각_주간	-2.8234	0.051	-55.264	0.000	-2.924	-2.723

- 추정된 로지스틱 회귀식을 바탕으로 인명피해발생 확률을 계산하여 지표 산출에 활용했다.
- 재산 피해액 - 다중선형 회귀분석
 - 인명피해발생여부 예측을 위해 사용한 독립변수 그대로 사용하여 재산피해액에 대해 모델링을 진행했다.
 - 종속변수로 사용된 재산피해액의 경우 편차가 심해 분산 안정화 변환 방법 중 하나인 로그 변환을 진행한 뒤 사용했다.

OLS Regression Results

Dep. Variable:	재산피해액	R-squared (uncentered):	0.784
Model:	OLS	Adj. R-squared (uncentered):	0.784
Method:	Least Squares	F-statistic:	7842.
Date:	Wed, 11 Oct 2023	Prob (F-statistic):	0.00
Time:	12:38:30	Log-Likelihood:	-46586.
No. Observations:	19484	AIC:	9.319e+04
Df Residuals:	19475	BIC:	9.326e+04
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
현장안전센터거리	0.1531	0.021	7.143	0.000	0.111	0.195
시간단위기온	-0.5016	0.037	-13.705	0.000	-0.573	-0.430
시간단위풍속	-0.0586	0.027	-2.175	0.030	-0.111	-0.006
시간단위가시거리	-0.3453	0.029	-12.005	0.000	-0.402	-0.289
실효습도	0.1372	0.030	4.566	0.000	0.078	0.196
가스사용량(KWh)	0.0146	0.002	7.867	0.000	0.011	0.018
전기사용량(KWh)	0.0146	0.002	7.867	0.000	0.011	0.018
화재발생시각_새벽	4.9860	0.049	101.420	0.000	4.890	5.082
화재발생시각_저녁	4.7796	0.044	108.744	0.000	4.693	4.866
화재발생시각_주간	5.0571	0.030	168.108	0.000	4.998	5.116

Omnibus:	1048.418	Durbin-Watson:	1.983
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1230.999
Skew:	0.587	Prob(JB):	4.92e-268
Kurtosis:	3.369	Cond. No.	4.38e+16

- 각 변수별 결과는 다음과 같으며, **MAE(Mean Absolute Error)는 2.06**을 기록했다.
- 모델의 설명력을 나타내는 지표인 **R-Squared의 값이 78.4%**로 높게 나와 설명력이 높은 모델임을 확인할 수 있었다.
- 추정된 선형회귀식을 바탕으로 예상 ln(재산피해액)을 계산하여 지표 산출에 활용했다.

결과

- 모델링 결과 인명피해 발생여부는 정확도 93.9% 재산 피해액은 R-Squared가 78.4%, MAE(Mean Absolute Error)는 2.06을 기록하며 준수한 성능을 보여주었다. 이러한 모델을 바탕으로 인명피해 발생 확률과 예상피해액 값을 추출했고, 화재피해등급을 도출하는 데에 활용했다.

<화재피해등급 도출>

- 재산 피해액 모델이 예측한 값에 MinMax Scaler를 적용하여 0~1 사이의 값(이하 재산피해지수)으로 표현했다.
- 사상자 발생 여부 모델은 0~1 사이의 확률값을 출력할 수 있다. 이 예측 확률에 MinMax Scaler를 적용하여 마찬가지로 0~1 사이의 값(이하 인명피해지수)으로 표현했다.
- 재산 피해액 지수와 인명피해지수를 곱한 값에 제곱근을 취하고 100을 곱한 값으로 얻은 화재피해지수를 통해 최종적으로 화재 피해 등급을 도출했다.

$$X_{\text{재산피해지수}} = \text{MinMax}(\hat{y}_{\text{예상피해액}})$$

$$X_{\text{인명피해지수}} = \text{MinMax}(\hat{y}_{\text{인명피해확률}})$$

$$Y_{\text{화재피해지수}} = \sqrt{X_{\text{재산피해지수}} * X_{\text{인명피해지수}}} * 100$$

$$(0 \leq Y_{\text{화재피해지수}} \leq 100)$$

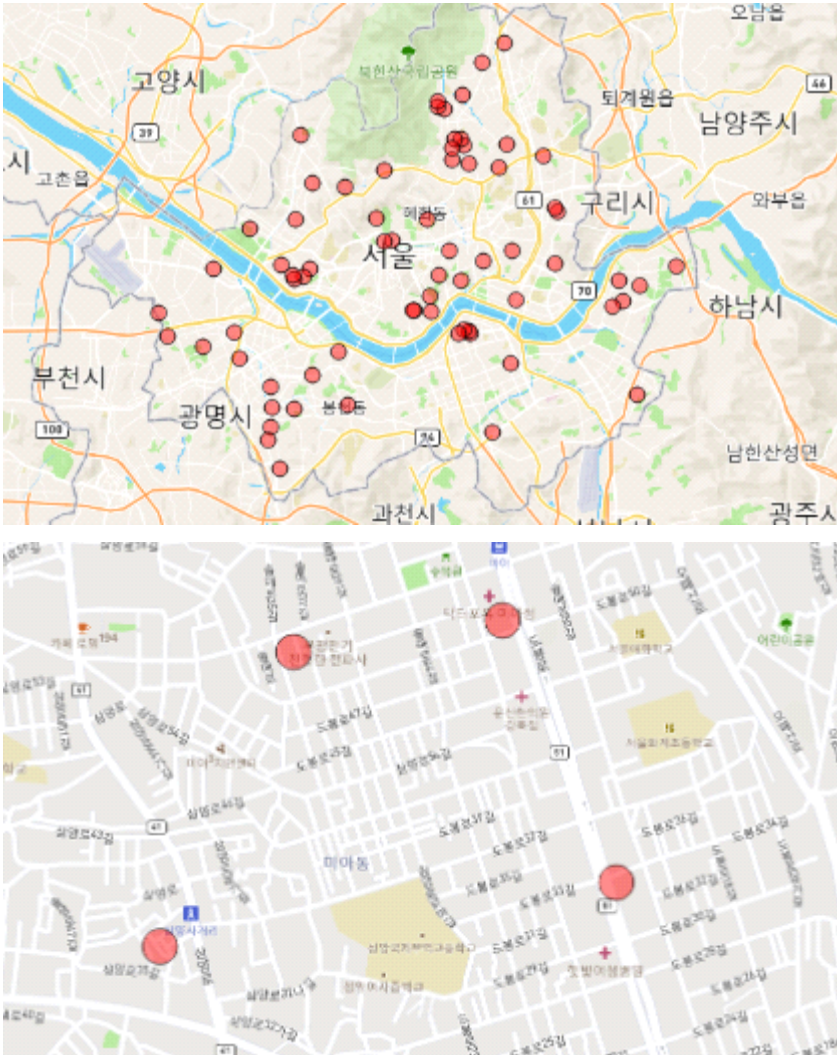
<최종결과>

화재피해등급	화재피해지수	건물 수
1등급	66 이상	70개
2등급	33 이상 66 미만	7,071개
3등급	33 미만	17,214개
총계	-	24,355개

- 피해등급에 따라 실시간으로 건물의 화재위험도를 파악하고 선제적 대응할 수 있다.

<1등급 건물 시각화>

- 태블로를 이용해 화재등급이 높은 건물을 시각화하여 직관적으로 확인할 수 있도록 제공한다.



5) 차별성 및 기대효과

건물 화재에 대한 화재 피해(인명피해, 재산 피해)를 예측하는 모델을 활용하여 '화재피해등급'을 산출할 수 있다. 화재피해등급의 변수 특성상 실시간 모니터링이 가능하며 건물 단위로 모델링을 진행하였기에 실시간 건물 단위 모니터링이 가능하다. 이를 활용한다면 건물 화재 발생에 대한 피해의 크기와 범위를 미리 파악할 수 있으며 화재 발생 시 건물 소유자, 시설 관리자, 정부 기관 등이 신속하고 선제적인 대처를 할 수 있기에 상황에 맞는 최적화된 대응 전략을 수립하여 화재 발생에 의한 피해를 최소화할 수 있을 것이다. 또한 정부에서 건물에 대한 효율적인 점검 스케줄링을 기획할 수 있어 많은 시간과 비용을 절약할 수 있을 것이다.

또한 추후 건물 이용 연령층, 건물 유동 인구 등 더 많은 데이터를 확보하여 다양한 변수를 고려한다면 모델의 정확도를 향상할 수 있다. 이를 통해 화재피해등급을 좀 더 세분화하는 것이 가능하며 중요한 의사결정 및 정책 등 활용 가능 범위를 넓힐 수 있을 것이다. 또한 데이터가 동적으로 수집되고 더 나은 지표를 개발함에 따라 모델 확장 및 개선이 용이하다. 이를 통해 계속 변화하는 상황 속에서 최적화된 대처 전략 모델을 바탕으로 화재피해등급 업데이트를 한다면 효과적인 대응 전략을 수립할 수 있을 것이다.

6) 융합데이터의 제안

건물 유동인구, 주 이용 연령층 등 화재 피해와 관련된 보다 더 다양한 데이터를 사용할 수 있으면 모델의 정확도가 좋아질 것이며, 각 데이터들을 실시간으로 수집할 수 있는 API를 구축한다면 '실시간 대응 전략'이라는 본질에 맞는 분석이 가능할 것이다.