

빅데이터 기반 건축물 화재 예측 모델 개발 연구

A Study on the Development of Building Fire Prediction Model Based on Big Data Analysis

조영진 Cho, Youngjin

허한결 Heo, Hankyul

안의순 Ahn, Euisoon

류수연 Ryu, Suyoon

송유미 Song, Yumi

현태환 Hyeon, Taehwan

(a u r i

일반연구보고서 2022-5

빅데이터 기반 건축물 화재 예측 모델 개발 연구

A Study on the Development of Building Fire Prediction Model Based on Big Data Analysis

지은이 조영진, 허한결, 안의순, 류수연, 송유미, 현대환
펴낸곳 건축공간연구원
출판등록 제2015-41호 (등록일 '08. 02. 18.)
인쇄 2022년 12월 26일, 발행: 2022년 12월 31일
주소 세종특별자치시 가림로 143, 8층
전화 044-417-9600
팩스 044-417-9608

<http://www.auri.re.kr>

가격: 10,000원, ISBN: 979-11-5659-402-4

이 연구보고서의 내용은 건축공간연구원의 자체 연구물로서 정부의 정책이나 견해와 다를 수 있습니다.

| 연구책임

조영진 연구위원

| 연구진

허한결 부연구위원

안의순 부연구위원

류수연 연구원

송유미 연구원

현태환 연구원

| 외부연구진

한국과학기술원 임리사 교수

백대화, 이정원, 김유진, 노재창 연구원

| 연구심의위원

유광흠 선임연구위원

조상규 선임연구위원

이여경 연구위원

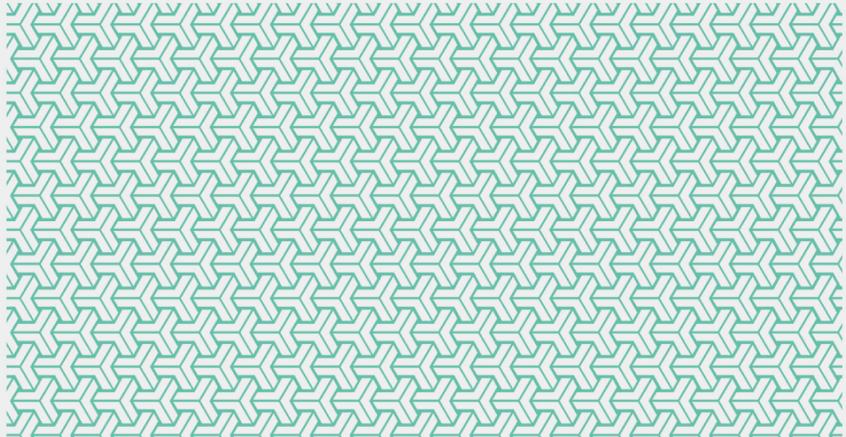
민범식 한아도시연구소 부회장

이강 연세대학교 교수

임은성 국토연구원 선임연구위원

연구요약

Summary



행정안전부에서 발간한 ‘2020 재해연보’에 따르면 2011년~2020년간 태풍, 호우, 지진으로 인한 건물 피해액이 매우 높게 나타났으며, 건축물에서 발생하는 지진, 홍수, 태풍 등 자연재난과 화재 및 범죄 등 사회재난 등으로 인한 국민의 불안감이 높아지고 있다. 특히, 건축물에서 발생하는 각종 사고 중 가장 많이 나타나는 것은 범죄이며, 그 다음은 화재이다. 범죄의 경우 발생 원인이 건축물인 경우보다는 범죄자 개인의 문제인 경우가 다수이기 때문에 사실상 우리가 경험하는 건축물이 원인인 재난은 화재가 가장 많은 것으로 볼 수 있다. 우리나라 건축물은 지속적으로 대형화/노후화 추세에 있어 화재 발생 시 대형 사고로 확대될 가능성이 높고, 노후 건축물의 증가는 건축물의 기계 및 전기 설비 노후화에 따른 화재 발생 위험을 증대시키는 요인으로 나타나고 있다.

최근 건축물의 재난·재해 문제 해결을 위해 건축·도시공간의 데이터 활용을 통하여 대응하고자 하는 시도가 활발히 이루어지고 있다. 화재안전 분야에서도 소방서의 위치 선정 등을 출동시간 데이터를 활용하고자 하는 시도 등 데이터 기반 재난 대응에 노력하고 있다. 다만 아직까지는 건축물 화재 대응은 화재확산방지 시설의 설치·점검 등을 중심으로 예방보다는 피해최소화에 초점을 맞추고 있는 실정이다. 사실상 모든 재난이 동일하지만 사고는 발생 후 피해를 최소화하는 것도 중요하지만 가장 좋은 것은 사고 자체를 사전에 예방하는 것이다.

AI, 머신러닝, 딥러닝 등 제4차 산업혁명 기술의 발달은 건축물에서 발생하는 재난·재해 발생과 피해 예측의 가능성을 제시하고 있으며, 이러한 기술을 활용하여 건축물의 안전사고를 사전에 예측하고자 하는 시도가 필요하다. 특히, 건축행정 시스템 및 국가화재정보시스템 등 건축물의 화재와 관련하여 다양하고 방대한 빅데이터가 공공데이터로 구축 및 개방되고 있는 만큼, 이러한 빅데이터를 활용한 재난·재해 예측 시스템 구축을 위한 연구가 필요한 실정이다.

이에 본 연구는 건축물의 속성정보와 건축물에서 발생한 화재 발생 데이터를 활용하여 건축물의 안전사고 예방을 위한 재난·재해 위험지역 예측 방법론을 구축하고, 방법론을 토대로 건축물 화재 예측 모델을 제안하고자 한다.

2장에서는 국내외 재난·재해 빅데이터 시스템 및 빅데이터 분석 방법론과 관련 연구동향을 고찰하였다.

국가는 재난 및 재해 예방·관리·복구를 위해 다양한 위기관리 시스템을 구축하여 운영하고 있으며, 관련 시스템은 다양한 재난·재해 데이터를 기반으로 운영되고 있다. 국내 재난·재해 예방 및 대응 주무부처는 행정안전부 재난안전관리본부가 총괄하고 있으며, 재난·재해 예방 및 대응을 위해 대표적으로 국가재난

관리정보시스템 및 안전정보통합공개시스템을 운영하고 있다.

사회재난인 화재와 범죄는 개별적인 시스템을 구축하여 운영하고 있다. 화재는 국가화재정보시스템을 통해 지역별 화재발생위험도에 관한 정보를 제공하여 사전에 화재발생을 예방하고, 화재예방정책 수립 및 대응자료로 활용되고 있다. 경찰청에서 운영하는 지리적 프로파일링 시스템은 범죄예방을 위한 범죄위험도 측정을 주로 수행하고 있다. 국외의 화재발생 예측 시스템은 우선적으로 대규모 화재 발생 모니터링을 통한 화재 발생 및 확산 예측 시스템을 주를 이루고 있다. 대표적으로 미국 비영리 환경 단체인 CI에서 운영하는 FIRECAST와 미국 내무부에서 운영하는 Fire Danger Forecast가 있다.

국내외 재난·재해 관련 빅데이터 예측 시스템을 살펴본 결과 재난·재해 예측을 통한 예방을 위해서는 관련 빅데이터의 수집·가공 및 분석을 위한 일련의 데이터 생태계가 필요하다는 것을 확인할 수 있다. 또한 재난·재해 발생에 대한 신속한 대응 및 조치와 피해최소화를 위해 정책수행자 및 일반국민이 이해하기 쉽게 관련 자료를 제공해야 하며, 건축물의 재난·재해 예측모델 구축을 위해서는 건축물 속성정보가 중요하다는 시사점이 있다.

빅데이터의 분석 방법론과 관련 연구동향을 보면 분석 대상과 목적에 따라 다양한 빅데이터 분석 방법이 개발 및 활용되고 있는 것을 확인할 수 있다. 특히 선행 연구를 검토에서 화재위험지역 예측모델 구축을 위해서는 분석 시나리오의 개발을 우선적으로 해야 하고, 시나리오 개발 시 수집·활용 빅데이터의 특성과 도출하고자하는 목적에 따라 모델의 유형을 달리 적용해야 한다는 결과가 도출되었다.

3장에서는 건축물 화재 예측을 위해 관련 데이터 선정 및 연계방안을 마련하고 화재발생 위험 건축물 예측 방법론을 수립하였다.

화재발생 위험 건축물 예측모델의 실증을 위하여 다양한 공간단위에 적용할 수 있도록 우리나라 도시의 범용적 특징을 가진 서울시를 연구 대상지로 선정하였다. 다음으로 빅데이터 분석 방법론에 적용하기 위해 본 연구에서 분석하고자 하는 화재데이터를 건축행정데이터에 연계 및 결합하였다. 결합방식은 화재발생 데이터와 건축행정 데이터의 구조를 확인하고, 두 데이터의 연계 KEY로 주소정보를 담고 있는 PNU 코드를 생성하여 연계를 수행하였다. 거리변수 적용을 위하여 연계데이터의 공간정보화 즉 지오코딩을 진행하였다. 지오코딩 방식은 건물 통합정보를 BaseMap으로 진행하였으며, 범죄데이터의 경우 자료가 방대하기 때문에 범죄발생 건축물을 추출하는 대신 집계구 단위로 가공하였다. 다음단계로 ArcGis프로그램을 활용하여 분석용 최종 데이터셋 구축을 위해 하였으며, 화

재관련 변수 3종, 범죄관련 변수 6종, 건축물 관련 변수 6종을 독립변수로 구축하고 건축물 화재발생 여부를 종속변수로 구축하였다.

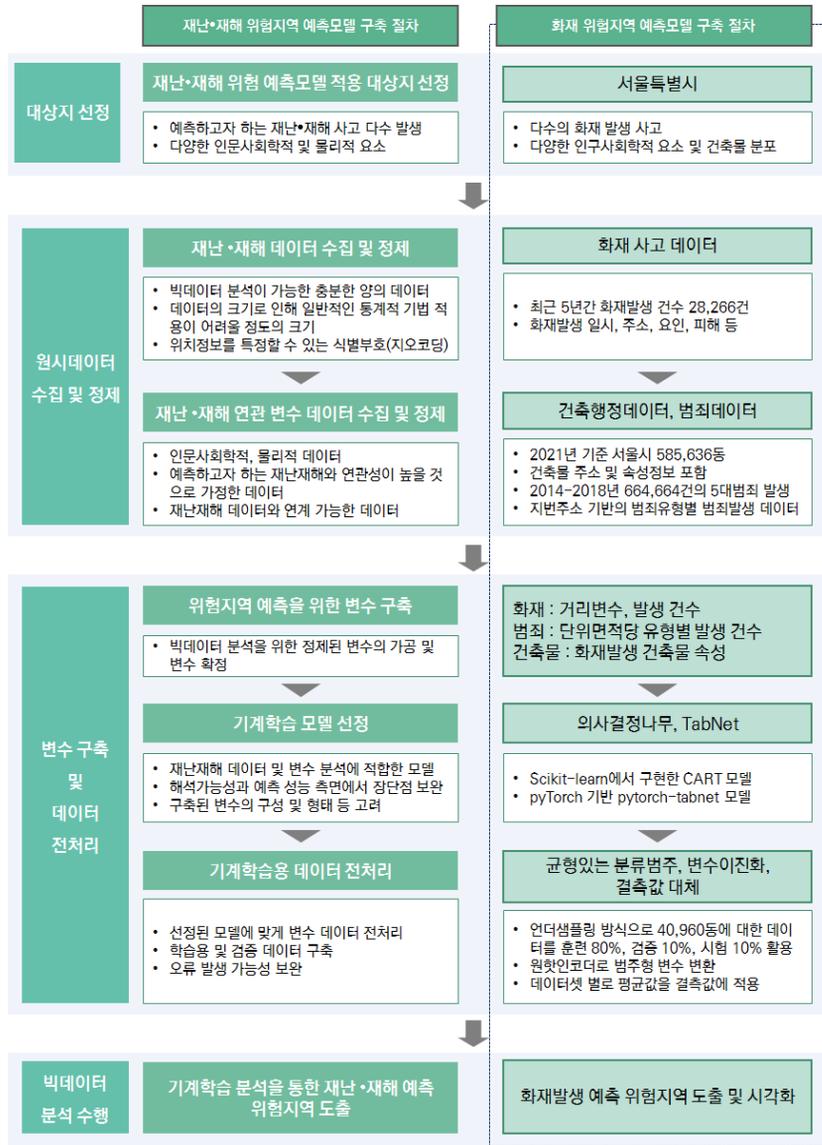
빅데이터 분석을 위한 변수를 설정한 후 기계학습모델 중 결과에 대한 설명이 가능한 의사결정나무와 특정 건축물의 화재발생 위험을 측정할 수 있는 딥러닝 모델을 분석모델로 선정하였다. 세부적으로는 의사결정나무의 여러 모델 중 화재 발생 여부가 이진변수(발생, 미발생)인 점을 고려하여 scikit-learn에서 구현한 CART모델을 사용하였다. 또한 딥러닝 모델 중에는 테이블 형태의 데이터를 간편하게 분석할 수 있는 pyTorch 기반 pytorch-tabnet 모델을 선정하였다.

이후 각 모델에 맞추어 전처리 및 언더샘플링을 수행하여 총 40,960동의 데이터를 최종 분석 데이터로 추출하였으며, 분석 대상 데이터를 무작위 표본 추출로 분할하여 80%는 훈련데이터로 활용하고, 10%는 검증데이터, 10%는 시험데이터로 활용하였다. 이와 같이 재난·재해 위험지역 예측 모델 개발은 대상지 선정-분석을 위한 데이터 수집-변수 구축을 위한 데이터 가공-변수구축-빅데이터 분석 모델 선정-모델에 적합하도록 변수 전처리-분석을 통한 예측모델 개발의 순으로 진행하였다.

4장에서는 3장에서 수행한 최종데이터셋 구축 및 기계학습 모델 선정을 기반으로 화재발생 위험 건축물 예측 모델을 적용하였다.

우선 건축물 화재 데이터 기술통계를 통해 서울시 건축물 화재 현황을 파악하였다. 서울시의 경우 공동주택 및 단독주택에서의 화재 발생 건수가 높았으며, 연면적의 경우 200㎡ 초과 500㎡ 이하의 건축물에서 화재 발생 건수가 많았으며, 층수별 화재 통계에서는 5층 이하의 저층에서 화재발생 건수가 많았다. 또한 건축물의 노후도에 따라 화재발생 건수가 많은 것으로 나타났다.

빅데이터 분석에서는 3장에서 선정한 의사결정나무와 딥러닝 모델을 활용하여 화재 취약 건축물을 예측하고, 화재 취약 건축물 유형 및 화재 위험 영향 요인을 추출하였다. 이러한 과정을 거쳐 예측 모델을 검증하고 화재 취약 건축물의 특성을 해석하였다. 의사결정나무 모델을 통한 분석 결과 연면적 연면적 5,160㎡ 이상인 건축물 중 교육연구시설을 제외한 유형에서의 화재 발생률이 12.0%로 가장 높게 나타났다.

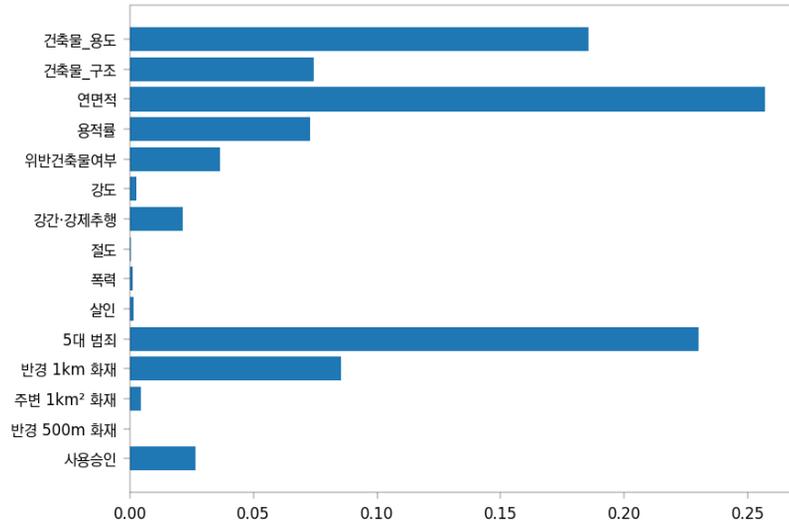


[그림] 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델

출처 : 연구진 작성

TabNet 모델에서는 구축한 화재 위험 영향 변수에 대한 딥러닝을 수행하여 예측 모델 활용의 가능성을 검증하였다. 분석 결과 검증 데이터셋에 대한 정확도는 67.5%로 나타났으며, 학습 및 조기 중단 판단에 사용하지 않은 시험 데이터셋에 대한 정확도는 65.8%로 나타났다. 화재 위험 영향 요인 추출 결과 건축물 속성 중에서는 연면적이 가장 연관성이 높게 나타났으며, 용도가 다음으로 연관성이 높게 나타났다. 위반건축물 여부 및 사용승인(노후도)는 상대적으로 연관성이 낮

게 나타났다. 범죄의 경우 5대범죄 건수가 연관성이 높게 나타난 반면, 개별 범죄 건수는 연관성이 낮게 나타났다. 거리변수의 경우 화재발생 지역 반경 1km가 주변 1km² 및 반경 500m에 비해 연관성이 높게 나타났다.



[그림] 건축물 화재 예측에서 TabNet 모델의 변수 중요도

출처 : 연구진 작성

본 연구에서 제안한 기계학습 기반 건축물 재난·재해 예측 모델은 개별 건축물 단위로 화재 취약성 예측, 해당 취약 유형, 취약 판단이 도출된 원인 등을 파악할 수 있는 방법론이다. 분석의 마지막 단계로서 화재 취약성 예측 결과를 공간정보 기반으로 시각화하여, GIS 시각화 기술을 적용한 활용 가능성을 검토하였다.

본 연구의 성과는 행정 빅데이터를 활용하여 건축물의 안전사고를 예측하는 방법론을 구축하고 화재를 대상으로 가능성을 확인했다는 것이다. 다만 화재 데이터 및 건축행정정보 데이터의 일부 오류로 인해 모든 데이터에 대한 분석을 수행하지 못한 점과, 빅데이터 분석에 있어 적용된 변수의 유형이 다양하지 못하여 최종적으로 예측 정확도가 상대적으로 낮다는 한계가 있다. 향후 '빅데이터 기반 건축물 안전사고 예측 모델'의 정확성 향상 및 고도화를 위해서는 관련 데이터의 무결성을 높이고, 다양한 인문·사회학적 변수를 적용할 필요가 있다.

주제어
빅데이터, 건축물화재, 의사결정나무, TabNet, 기계학습

차례 CONTENTS

제1장 서론	1
1. 연구의 배경 및 목적	2
1) 연구배경 및 필요성	2
2) 연구의 목적	6
2. 연구의 범위 및 방법	7
1) 연구의 범위	7
2) 연구의 방법	8
3. 선행연구와의 차별성	10
제2장 재난·재해 빅데이터 관련 동향	13
1. 국내 재난·재해 빅데이터 시스템	14
1) 국가재난관리정보시스템(NDMS)	14
2) 안전정보통합공개시스템	16
3) 국가화재정보시스템(NFDS)	20
4) 경찰청 지리적 프로파일링 시스템(GeoPros)	22
2. 국외 재난·재해 빅데이터 시스템	24
1) FIRECAST by Conservation International	24
2) Fire Danger Forecast by USGS(U.S. Department of the Interior)	25
3) 건축물 화재 위험도 예측 모델(Metro21: Smart Cities Institute)	28
4) 구글(Google) 홍수예측 시스템	30
3. 기계학습 모형 고찰	32
1) 의사결정나무	33
2) 랜덤포레스트	34
3) 서포트 벡터 머신	35
4) 인공신경망	36
5) 딥러닝	37

차례 CONTENTS

4. 빅데이터 기반 재난·재해 예측 연구	39
1) 빅데이터 기반 재난재해 연구 동향	39
2) 빅데이터 기반 화재 연구 동향	45
5. 소결	48
1) 국내외 재난·재해 빅데이터 기반 예측·예방·대응 시스템	48
2) 빅데이터 분석 방법론 및 관련 연구 동향	49
제3장 건축물 화재 예측 모델 개발	51
1. 대상지 선정 및 데이터 연계	52
1) 대상지 선정 및 특성	52
2) 데이터 연계 및 공간정보화 방향	57
3) 건축물 화재 발생 데이터 공간정보화	59
4) 건축물 범죄 발생 데이터 공간정보화	65
2. 건축물 화재 위험지역 예측 방법론 수립	67
1) 기계학습 기반의 화재 및 범죄 취약건물 예측모형 개발	67
2) 화재 및 범죄 데이터 가공	67
3) 분석 데이터 구축	72
4) 기계학습 모델 선정 및 분석 데이터 전처리	74
5) 기계학습을 활용한 화재 취약 건축물 예측	79
3. 소결	80
제4장 건축물 화재 예측 모델 적용	83
1. 건축물 특성에 따른 화재 통계	84
2. 기계학습 기반 화재 취약 건축물 예측	97
1) 의사결정나무 모델을 활용한 화재 취약 건축물 유형 추출	97
2) TabNet 모델을 활용한 화재 위험 영향 요인 추출	100

차례 CONTENTS

3) 화재 취약 건축물 예측 및 시각화	103
3. 소결 및 시사점	104
1) 건축물 화재 데이터 기술통계	104
2) 기계학습 기반 화재 취약 건축물 예측	105
제5장 결론	107
1. 연구의 요약	108
2. 향후 연구과제 및 정책적 시사점	110
참고문헌	113
SUMMARY	117

[표 1-1] 2011~2020년 자연재해로 인한 건물 피해현황	2
[표 1-2] 2012~2021년 연면적별 건축물 동수	3
[표 1-3] 주요 건축물 안전사고 발생현황	4
[표 1-4] 2016년 서울시 건축물 경과년수 별 화재발생 현황	5
[표 1-5] 2016년 서울시 건축물 연면적 별 화재발생 현황	5
[표 2-1] 국가재난관리정보시스템 시스템 구성 및 주요기능	14
[표 2-2] 안전정보통합공개시스템 2020년 공개대상 정보	16
[표 2-3] 생활안전지도 주제별 공개정보 내역	18
[표 2-4] 화재통계 제공 내역	20
[표 2-5] 화재 발화요인별 세부요인	21
[표 2-6] 화재통계에서의 화재장소 분류체계	22
[표 2-7] FIRECAST의 주요 서비스	25
[표 2-8] Fire Danger Forecast의 주요 서비스	27
[표 2-9] 데이터 매쉬업을 위한 활용 데이터	39
[표 2-10] 침수 대응 재해지도 구축 활용 데이터	41
[표 2-11] 빅데이터 활용 재난관리 시스템	42
[표 2-12] 재해 분야별 안전지수 모델 분석 결과	45
[표 2-13] 화재영향인자 데이터 평가 결과	46
[표 2-14] 화재현장 위험도의 정의 및 등급화	47
[표 3-1] 서울특별시 건축물 용도·면적·층수별 현황	54
[표 3-2] 서울시 용도지역별 면적 및 비율	55
[표 3-3] 2017~2021년 서울시 발화요인별 화재발생 건수	56
[표 3-4] 2017~2021년 서울시 장소별 화재발생 건수	56
[표 3-5] 화재데이터 지오코딩에 따른 오류구분 및 설명	60
[표 3-6] 화재데이터 지오코딩에 따른 오류구분 및 예시	60
[표 3-7] 화재데이터 Point Relocation 칼럼코드 및 설명	62
[표 3-8] 건축물 대상과 화재 데이터 매칭 결과	64
[표 3-9] 화재 건축물 연계 데이터 최종 매칭 결과	64
[표 3-10] 화재 데이터 및 건축물 대상 매칭 결과 칼럼Code 및 설명	65
[표 3-11] 범죄 데이터 지오코딩 성공률	66

표차례 LIST OF TABLES

[표 3-12] 범죄 유형별 면적당 범죄발생횟수 기술통계	70
[표 3-13] 화재발생 건축물 데이터셋의 화재 및 범죄관련 변수	72
[표 3-14] 화재발생 건축물 데이터셋의 화재 및 범죄관련 변수	74
[표 3-15] 원핫 인코더 적용 예시	78
[표 4-1] 최근 5년간(2017~2021년) 서울시 건축물 용도별 화재 발생 건수	84
[표 4-2] 최근 5년간(2017~2021년) 서울시 건축물 연면적별 화재 발생 건수	85
[표 4-3] 최근 5년간(2017~2021년) 서울시 건축물 층수별 화재 발생 건수	85
[표 4-4] 최근 5년간(2017~2021년) 서울시 건축물 노후도별 화재 발생 건수	86
[표 4-5] 서울시 구별 인구수/면적 대비 인명피해	89
[표 4-6] 지하층수에 따른 평균 피해 정도	93
[표 4-7] 건축물 유형별 따른 인명피해	94
[표 4-8] 건축물 사용승인 연도별 인명피해	95
[표 4-9] 건축물 사용승인 연도 및 용도 별 화재 인명피해	96
[표 4-10] 재난·재해 취약 건축물 유형별 화재 발생률 비교	99

그림차례 LIST OF FIGURES

[그림 1-1] 장소별-발화요인별 화재발생 현황(2020년)	5
[그림 1-2] 연구흐름도	9
[그림 2-1] 건축물 및 시설물 안전점검정보 지도 예시	18
[그림 2-2] 안전정보통합공개시스템의 생활안전지도 범죄주의구간 예시	19
[그림 2-3] 안전정보통합공개시스템의 생활안전지도 노후건물정보 예시	19
[그림 2-4] GeoPros 범죄위험지역(Hot Spot) 분석 예시	23
[그림 2-5] Active Fire Live Map	24
[그림 2-6] Fire Danger Forecast에서 제공하는 최근 7일 동안의 WFPI 지도	26
[그림 2-7] Fire Danger Viewer Application	26
[그림 2-8] 건축물 유형에 따른 예측 정보를 보여주는 데이터 대시보드	29
[그림 2-9] 고위험 건축물을 보여주는 인터랙티브 맵	29
[그림 2-10] 구글 홍수 알림 대시보드	30
[그림 2-11] 구글 홍수 예측 모바일 앱	31
[그림 2-12] 머신러닝의 주요 활용분야	32
[그림 2-13] 의사결정나무의 일반적 형태	33
[그림 2-14] 부트스트랩 샘플링 개념	34
[그림 2-15] 랜덤포레스트 기반의 앙상블 분류모델 구축 개념	35
[그림 2-16] 서포트 벡터 머신의 기본 개념	36
[그림 2-17] 뉴런을 모방한 퍼셉트론의 개념	36
[그림 2-18] 단층 퍼셉트론과 다층 퍼셉트론	37
[그림 2-19] 학습을 위해 필요한 데이터량과 모형의 성능사이 관계	38
[그림 2-20] 데이터 매쉬업을 통한 재난안전 유형 판별 알고리즘	39
[그림 2-21] 재난안전 맵 활용(예)	40
[그림 2-22] 수요자 중심의 안전도시 에코시스템(예)	40
[그림 2-23] 빅데이터 스키마의 구축 및 적용	41
[그림 2-24] 분산병렬 재난 위험도 분석 처리 흐름도	42
[그림 2-25] 필요 정보 추출 시스템의 흐름	43
[그림 2-26] 호우 및 홍수로 인한 산사태 발생 시나리오	44
[그림 2-27] 안전지수 서비스 모델 구조	44
[그림 2-28] 이용자 위치기반 중심 실시간 데이터 반영 안전지수 표출 시스템 설계도	45
[그림 2-29] 화재예측플랫폼 예시	46

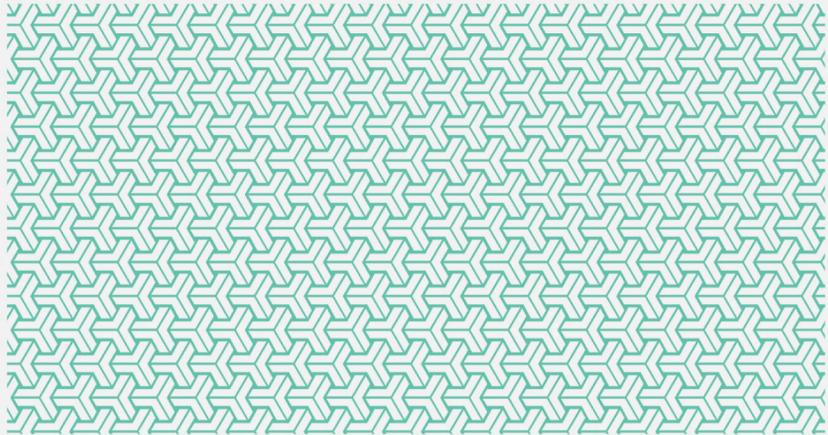
그림차례 LIST OF FIGURES

[그림 2-30] 머신러닝 알고리즘 분석 결과	47
[그림 3-1] 대상지 선정기준	53
[그림 3-2] 건축물 데이터와 화재 데이터 연계 절차	57
[그림 3-3] 건축물 화재 발생 데이터 제공 정보 일부 예시	58
[그림 3-4] 건축물 화재 발생 데이터 구축 프레임워크	59
[그림 3-5] QGIS Plug-in web_geocoder 활용	60
[그림 3-6] 화재 데이터의 공간 정보화 완료	62
[그림 3-7] 건축물범죄 발생 데이터 구축 프레임워크	66
[그림 3-8] 기계학습 기반의 화재 취약건물 예측모형 개발 흐름	67
[그림 3-9] 범죄데이터 공간화	68
[그림 3-10] 집계구와 지번주소 경계(지번주소: 붉은색, 집계구: 파란색)	69
[그림 3-11] 집계구별 범죄 발생률 지도	69
[그림 3-12] 공간자기상관의 예시	70
[그림 3-13] 가장 근접한 화재 발생 포인트 검색 및 값 부여 개념	71
[그림 3-14] 화재발생 밀도 분석 및 공간화 결과	71
[그림 3-15] 가장 근접한 화재 발생 포인트까지의 KDE 분석 결과 그래프	73
[그림 3-16] scikit-learn의 의사결정나무(CART) 모델 분석 결과 예시	75
[그림 3-17] TabNet 모델의 신경망 구조	76
[그림 3-18] pytorch-tabnet의 GitHub 페이지	76
[그림 3-19] 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델	82
[그림 4-1] 서울시 구별 화재 사망자 수 합계	86
[그림 4-2] 서울시 구별 화재 부상자 수 합계	87
[그림 4-3] 서울시 구별 인명피해 합계	87
[그림 4-4] 서울시 구별 재산피해 합계	88
[그림 4-5] 서울시 구별 인구수 대비 인명피해	88
[그림 4-6] 서울시 구별 면적 대비 인명피해	88
[그림 4-7] 지상 층수에 따른 합계 사망자 수	90
[그림 4-8] 지상 층수에 따른 합계 부상자 수	90
[그림 4-9] 지상 층수에 따른 합계 화재 인명피해	90
[그림 4-10] 지상 층수에 따른 합계 화재 재산피해	90

그림차례 LIST OF FIGURES

[그림 4-11] 지상 층수에 따른 평균 사망자 수 히스토그램	91
[그림 4-12] 지상 층수에 따른 평균 부상자 수 히스토그램	91
[그림 4-13] 지상 층수에 따른 평균 인명피해 히스토그램	91
[그림 4-14] 지상 층수에 따른 평균 재산피해 히스토그램	92
[그림 4-15] 지하 층수에 따른 평균 사망자 피해	92
[그림 4-16] 지하 층수에 따른 평균 부상자 피해	92
[그림 4-17] 지하 층수에 따른 평균 인명피해	93
[그림 4-18] 건축물 용도별 화재 인명피해 히스토그램	94
[그림 4-19] 건축물 사용승인 연도별 화재 인명피해	95
[그림 4-20] 건축물 사용승인 연도 및 시설종류에 따른 화재 인명피해	95
[그림 4-21] 건물 지상층수와 건물 승인년도에 따른 화재 인명피해	96
[그림 4-22] 의사결정나무 모델 규모에 따른 훈련 및 검증 데이터셋 예측 정확도	97
[그림 4-23] 건축물 화재 발생 예측 의사결정나무 모델	98
[그림 4-24] TabNet 모형 학습 과정	100
[그림 4-25] 건축물 화재 예측에서 TabNet 모델의 변수 중요도	101
[그림 4-26] TabNet 모델의 각 사례별 종합 및 단계별 변수 중요도	102
[그림 4-27] 건축물 화재 발생가능성 예측 결과	103

제1장 서론



1. 연구의 배경 및 목적
2. 연구의 범위 및 방법
3. 선행연구와의 차별성

1. 연구의 배경 및 목적

1) 연구배경 및 필요성

□ 기후 및 사회 변화에 따른 건축물의 대형 안전사고 증대

기후변화에 따른 홍수, 태풍, 지진 등 자연재난과 사회적 여건 변화에 따른 화재, 범죄, 붕괴 등의 사회재난으로 인한 안전사고가 다발하고 있다. 특히 점차 대형화되는 건축물의 재고에 따라 대규모 인명과 재산 피해에 대한 불안감이 높은 실정이다. 이러한 건축물의 재난·재해로 인한 피해를 방지하고자 정부는 정책 및 제도를 통해 피해 최소화 및 예방방안을 마련하고 있으나, 급변하는 사회변화에 선제적으로 대응하기에는 명확한 근거가 부족한 실정이다. 특히 건축물 및 사회간접자본이 밀집되어 있는 도시지역의 경우 재난·재해가 발생할 경우 피해규모는 상상할 수 없을 정도로 클 것으로 예상되고 있다.

행정안전부에서 발간한 ‘2020 재해연보’의 자연재해로 인한 피해현황을 살펴보면, 건물의 경우 태풍, 호우, 지진으로 인한 피해액이 매우 높은 것으로 나타났다. 다만 1회성 재해로 인한 피해는 지진은 발생 시 그 피해가 막대한 것으로 나타났다. 우리나라는 최근 10년간 총 1,040회(북한포함)의 지진이 발생하였으며, 특히 경북지역은 460회로 전체 지진발생의 44.2%를 차지하고 있어 주의가 요구된다. 대표적으로 2017년 포항시에서 발생한 지진으로 인해 많은 건축물이 휘어지거나 붕괴되었고, 이로 인한 부상자도 다수 발생하였다.

[표 1-1] 2011~2020년 자연재해로 인한 건물 피해현황

(금액단위 : 백만원)

일시	태풍	호우	대설	강풍	풍랑·강풍	태풍·호우	지진
2011	4,105	22,540	779	-	-	-	-
2012	29,015	2,955	-	334	-	-	-
2013		1,684	15	15	-	-	-
2014	109	3,144	371	1	-	-	-
2015	4	226	-	32	-	-	-
2016	3,089	1,345	94	-	355	-	4,508
2017		6,664	-	-	18	-	58,884
2018	2,179	3,702	-	-	32	434	-
2019	9,780	353	-	-	22	-	-
2020	6,133	31,238	-	-	43	-	-
합계	54,414	73,852	1,259	382	469	434	63,392

출처 : 행정안전부(2021).2020 재해연보, pp.263-273. 연구진 재구성

□ 중대형 건축물 증대로 인한 안전사고 발생 시 대형 인명 및 재산피해 우려

2012~2021년 건축물 현황통계에 따른 중대형 건축물(연면적 3천㎡이상) 현황을 살펴보면, 연면적 3천㎡~1만㎡의 건축물은 2012년 124,836동에서 2021년 165,150동으로 40,314동이 증가(32.3%)하였으며, 연면적 1만㎡이상 건축물은 47,530동에서 65,567동으로 18,037동이 증가(37.9%) 증가하였다. 전체 건축물에서 차지하는 비율도 2012년 2.5%에서 2021년 3.2%로 꾸준히 증가하고 있는 추세를 보이고 있다.

[표 1-2] 2012~2021년 연면적별 건축물 동수

연도	연면적	동수	비율	연도	연면적	동수	비율
2012	전체건축물	6,796,239	100	2017	전체건축물	7,126,526	100
	1천㎡~3천㎡	229,884	3.4		1천㎡~3천㎡	261,841	3.7
	3천㎡~1만㎡	124,836	1.8		3천㎡~1만㎡	148,306	2.1
	1만㎡이상	47,530	0.7		1만㎡이상	56,194	0.8
	합계	402,250	5.9		합계	466,341	6.5
2013	전체건축물	6,851,802	100	2018	전체건축물	7,191,912	100
	1천㎡~3천㎡	235,976	3.4		1천㎡~3천㎡	270,159	3.8
	3천㎡~1만㎡	128,722	1.9		3천㎡~1만㎡	153,346	2.1
	1만㎡이상	48,868	0.7		1만㎡이상	58,924	0.8
	합계	413,566	6.0		합계	482,429	6.7
2014	전체건축물	6,911,288	100	2019	전체건축물	7,243,472	100
	1천㎡~3천㎡	241,608	3.5		1천㎡~3천㎡	278,544	3.8
	3천㎡~1만㎡	132,811	1.9		3천㎡~1만㎡	158,115	2.2
	1만㎡이상	50,146	0.7		1만㎡이상	61,495	0.8
	합계	424,565	6.1		합계	498,154	6.9
2015	전체건축물	6,986,913	100	2020	전체건축물	7,275,266	100
	1천㎡~3천㎡	248,442	3.6		1천㎡~3천㎡	286,094	3.9
	3천㎡~1만㎡	137,980	2.0		3천㎡~1만㎡	162,179	2.2
	1만㎡이상	52,109	0.7		1만㎡이상	63,701	0.9
	합계	438,531	6.3		합계	511,974	7.0
2016	전체건축물	7,054,733	100	2021	전체건축물	7,314,264	100
	1천㎡~3천㎡	254,699	3.6		1천㎡~3천㎡	289,370	4.0
	3천㎡~1만㎡	143,036	2.0		3천㎡~1만㎡	165,150	2.3
	1만㎡이상	53,878	0.8		1만㎡이상	65,567	0.9
	합계	451,613	6.4		합계	520,087	7.1

출처 : KOSIS 국가통계포털. 면적별 건축물 현황 통계 재구성, https://kosis.kr/statHtml/statHtml.do?orgId=116&tblId=DT_MLTM_540&vw_cd=MT_ZTITLE&list_id=M1_5&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=M_T_ZTITLE (검색일 : 2022.03.21.)

이처럼 중대형 건축물이 증가로 지진 및 화재 등 안전사고 발생 시 대형 인명피해에 대한 우려가 크다. 특히 최근 빈번하게 발생하는 건축물 붕괴 및 화재 사고 등 건축물 안전에 대한 국민의 관심이 높아지고 있다. 또한 공사 중인 건축물의 안전 사고에 대한 불안감도 높은 실정이다. 건설공사 안전관리 종합정보망에 따르면 2021년 건설현장에서 발생한 붕괴, 전도, 화재 등 안전사고가 총 34,385건¹⁾으로 집계되었다.

건축물에서 발생한 주요 안전사고를 살펴보면 건축물 공사 중 또는 해체 중에 발생한 사고, 건축물 사용 중에 다중이용시설 및 대형건축물에서 발생하는 사고 등 건축물 전 생애에 걸쳐 여러 유형의 안전사고가 발생하는 것을 확인할 수 있다.

[표 1-3] 주요 건축물 안전사고 발생현황

일자	사고건축물	사고유형	피해
2017.12.21.	제천 스포츠센터	화재	사망 29명, 부상 37명
2018.01.26.	밀양 세종병원	화재	사망 50명, 부상 142명
2019.09.24.	김포 요양병원	화재	사망 2명, 부상 47명
2020.04.29.	이천 물류창고	화재	사망 38명, 부상 10명
2021.04.04.	광주 주택 리모델링	붕괴	사망 2명, 부상 2명
2021.06.09.	광주 학동 철거건물	붕괴	사망 9명, 부상 8명
2021.06.17.	이천 쿠팡물류센터	화재	사망 1명, 부상 1명

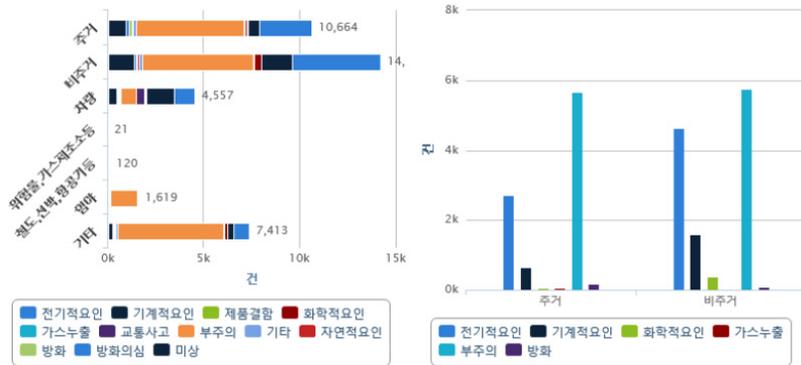
출처 : 최명영(2018). 반복되는 대형화재의 시사점과 위험관리 개선방안, 화재보험협회
 이영진(2019). '김포 요양병원 화재' 2명 사망·47명 부상, 합동감식, KBS News(2019.09.24.)
 김보형(2020). 2020년 주요 안전사고 총정리, 안전저널(2020.12.18.)
 김보형(2021). 2021년 주요 안전사고·이슈 총정리, 안전저널(2021.12.30.)

□ 노후 건축물 증가로 기계 및 전기 설비 노후화 등 화재발생 위험 증대

우리나라 노후 건축물은 매해 증가하고 있으며, 이러한 노후 건축물의 증가는 설비와 구조 노후화로 인해 화재발생 위험이 높은 것으로 나타났다. 2020년 소방청 화재통계연감에 따르면 화재발생요인 중 실화(부주의)를 제외하고 건축물에서의 화재 요인은 '전기적 요인' 및 '기계적 요인'으로 인한 화재발생 건수가 높았으며, 특히 비주거용 건축물의 경우 전기적 요인에 의한 화재발생 비율이 매우 높았다. 이는 건축물의 노후화로 인한 합선, 설비장비의 노후화로 화재가 발생하였다는 것을 의미한다. 또한 비주거용 건축물에서의 전기적·기계적 요인으로 인한

1) 건설공사 안전관리 종합정보망, <https://www.csi.go.kr/hrp/stats/hrpStatsByAccdList.do>, (검색일 : 2022.02.24.)

화재 발생건수가 주거용에 비해 높은 것은 체계적인 유지관리가 이루어지지 않는다는 것을 의미하며, 이에 대한 대책이 필요하다.



[그림 1-1] 장소별-발화요인별 화재발생 현황(2020년)

출처 : 소방청(2021), 2020년도 화재통계연감, p.77.

2016년 서울시 건축물 화재발생 데이터를 분석한 결과, 20년 이상 노후화된 건축물의 화재 발생건수는 총 2,529건으로 전체 건축물 화재발생 대비 약 59.7%로 나타났으며, 연면적 1,000㎡미만 소규모 건축물에서의 화재발생 건수는 2,767건으로 전체 건축물 화재발생 대비 약 65.3%로 나타났다. 이는 연면적 1,000㎡이하 건축물의 수가 중대형 건축물의 수보다 많기 때문인 것이기도 하지만, 그에 따른 발생 건수도 높아 소규모 노후 건축물에 대한 화재안전 확보를 위해 관리가 필요하다는 것을 시사하고 있다.

[표 1-4] 2016년 서울시 건축물 경과년수 별 화재발생 현황

경과년수		10년 미만	10~15년	15~20년	20~25년	25~30년	30~35년	35년 이상	기타	합계
		화재 발생 건수	471	444	408	737	869	474	749	84
화재 발생 현황	비율	11.1%	10.5%	9.6%	17.4%	20.5%	11.2%	17.7%	2.0%	100%

출처 : 건축공간연구원(2021), 2021년 건축물관리지원센터 업무 위탁 보고서, p.219.

[표 1-5] 2016년 서울시 건축물 연면적 별 화재발생 현황

연면적(㎡)		200 미만	200~300	300~500	500~1,000	1,000~3,000	3,000~10,000	10,000 이상	합계
		화재 발생 건수	676	496	691	904	659	389	421
화재 발생 현황	비율	16.0%	11.7%	16.3%	21.3%	15.6%	9.2%	9.9%	100%

출처 : 건축공간연구원(2021), 2021년 건축물관리지원센터 업무 위탁 보고서, p.219.

□ 건축물 재난·재해 문제 해결을 위한 건축·도시공간 데이터 활용 연구 필요

건축물에서의 재난·재해 및 화재 등 안전사고의 통상 대응 절차는 건축물 재난·재해 발생 시 소방서·경찰서 등 대응 부서에서 사건·사고인지 후 출동하며, 육안으로 건축물 외관 확인 후 현장 진입하여 구조 등을 진행하고 있다. 이때 필요시 건축물의 도면을 건축행정시스템(세움터)에 요청하여 이를 활용하고 있으나, 화재·공법·관리정보 등 건축물의 특성을 사전에 파악할 수 없어 현장인력의 경험에 의존하여 대응하고 있는 실정이다. 데이터 품질 개선 및 머신러닝 등의 데이터 처리 기술 발전으로 정부는 부분적으로 건축물 데이터를 활용한 안전사고 대응을 위해 노력하고 있다. 다만 관련 데이터 활용에 대한 체계적인 방법론은 여전히 부족한 실정이다.

건축물에 관한 정보는 건축행정시스템(세움터) 등 여러 공공시스템이 운영 중에 있으며, 이러한 시스템을 통한 건축물의 속성정보가 지속적으로 구축되고 있다. 또한 국가화재정보시스템 등 건축행정정보를 토대로 하는 안전관련 2차데이터의 생산 증가로, 활용 가능한 건축·도시공간 정보량 증가하고 있다. AI, 머신러닝, 딥러닝 등 4차산업혁명 기술의 발달로, 재난·재해 발생과 피해 예측 정확도가 현저히 증가하였으며, 건축·도시공간 데이터를 활용한 문제해결 능력 확보하고 있기 때문에 이러한 능력을 적용하기 위한 연구가 필요하다.

2) 연구의 목적

본 연구의 목적은 건축물의 속성정보와 건축물에서 발생한 화재 발생 데이터를 활용하여 건축물의 안전사고 예방을 위한 빅데이터 기반 재난·재해 위험지역 예측 방법론을 구축하고, 방법론을 토대로 건축물 화재 예측 모델을 개발하고 이를 실증하는 것이다. 연구의 세부적인 목적은 다음과 같다.

첫째, 건축물 화재발생 데이터와 건축행정 빅데이터의 연계 방안을 도출하고, 이를 선행연구와 사례 등과 종합하여 빅데이터 기반 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델을 개발한다.

둘째, 개발한 건축물 화재 예측 모델을 대상지에 적용하고 모델의 타당성을 검증한다. 검증 결과를 바탕으로 연구 개발한 빅데이터 기반 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델의 정책적 활용 방향을 제안한다.

2. 연구의 범위 및 방법

1) 연구의 범위

□ 시·공간 범위

본 연구에서는 건축물 속성정보 데이터와 건축물 화재 데이터를 결합하여 안전 사고 위험이 높은 건축물을 도출하는 것을 목적으로 하고 있다. 이에 다양한 유형의 건축물 및 안전사고 데이터를 구축하고 있는 서울시를 본 연구의 공간적 범위로 설정하며, 화재 데이터는 최근 5년(2017~2021년)간 발생데이터를 기준으로 하였다. 분석 대상 건축물의 유형은 모든 유형의 건축물을 분석대상으로 하되, 피해저감 전략 및 정책적 시사점 도출에 있어 필요한 유형의 경우 세부적으로 구분하여 분석하고자 한다.

□ 내용적 범위

본 연구는 최종적으로 화재 데이터와 건축행정 데이터의 연계방안을 도출하고, 이를 토대로 선행연구와 사례 분석을 종합하여 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델을 개발하고자 한다. 이에 따른 본 연구의 내용적 범위는 크게 화재 데이터와 건축행정 데이터의 연계, 연계 빅데이터의 기계학습을 위한 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델 개발, 마지막으로 건축물 화재 예측 모델 실증으로 구분된다.

- 건축행정 데이터와 화재발생 데이터 연계방안 도출

건축행정 정보시스템은 건축물의 행정처리를 위한 시스템으로 건축물 속성에 대한 다양한 정보를 구축하고 있다. 다만 근래까지 행정처리의 효율성 측면에서 시스템이 구축·운영 되어왔기 때문에 데이터로서 활용 측면에서는 미흡한 점이 있으며, 수작업으로 각종 속성 및 항목을 입력하기 때문에 일부 데이터의 오류가 있다. 이에 건축물 속성 데이터와 화재 데이터의 결합은 기본적으로 주소정보를 통해 이루어지고, 일부 매칭이 되지 않는 데이터의 경우 도로명주소시스템 등을 통해 확인작업을 거쳐 데이터 연계방안을 도출하고자 한다. 데이터 연계 이후에는 이를 공간정보화하여 대상지 화재 발생 현황을 시각적으로 표현하고자 한다.

- 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델 개발

건축물 속성 정보와 건축물 화재발생 정보를 연계한 빅데이터의 구축과정과 선

행연구 및 사례분석을 토대로 일련의 기계학습 시나리오를 구축하고, 구축한 시나리오를 기반으로 재난·재해 위험지역 예측 방법론을 수립하고, 수립한 예측 방법론을 활용하여 건축물 화재 예측 모델을 개발한다. 대상지의 건축물 속성과 화재 발생 속성을 연계한 빅데이터의 분석을 위해 머신러닝 및 다양한 빅데이터 분석 방법론을 활용할 예정이며, 빅데이터를 분석하는 것이기 때문에 분석 시나리오를 설정하는 것이 본 연구에서 중요한 부분이라 할 수 있다.

- 건축물 화재 예측 모델 실증 및 정책적 활용 방안 제안

개발한 건축물 화재 예측 모델의 타당성 검증을 위하여 실제 대상지에서 기계학습 기반 건축물 화재 발생을 예측한다. 이를 통해 화재 발생 위험 건축물의 특성을 도출하고, 화재 발생 위험이 높은 건축물을 도출한다. 화재 발생 위험이 높은 건축물이 도출한 이후에는 도출 지역의 공간적 특성을 분석하고, 이를 기반으로 건축물, 지역 및 공간의 재난·재해를 예측하는 방법론을 실증하여, 이를 정책적으로 적용하기 위한 시사점을 제안하고자 한다.

2) 연구의 방법

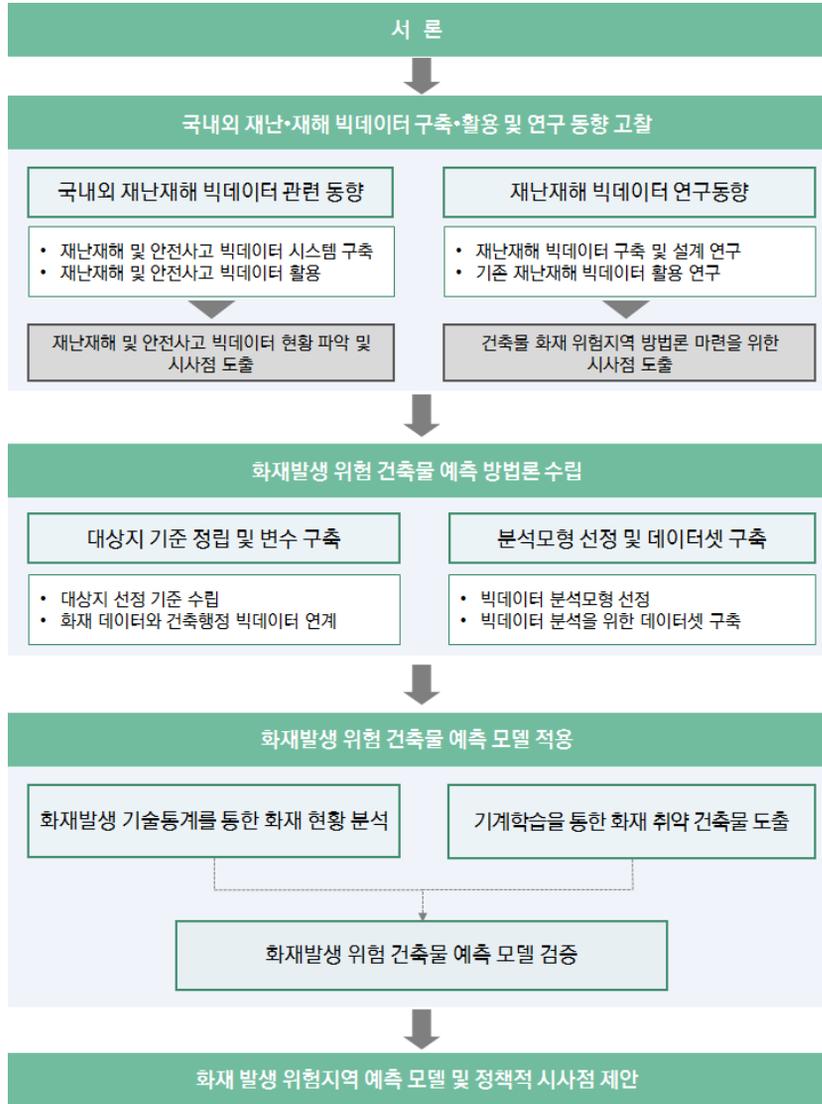
□ 문헌조사 및 전문가 활용

건축 및 도시공간의 물리적 속성을 중심으로 화재와 연관된 국내외 문헌을 수집 및 분석하고, 빅데이터를 활용한 건축 및 도시공간을 분석한 국내외 문헌에 대한 고찰을 수행한다. 이를 통하여 건축 및 도시공간의 안전에 대한 각종 지표 및 지수 산출 방법과 다양한 안전사고 발생 요인을 도출하고, 각종 이기종 데이터의 연계방법론 도출과 빅데이터 분석 모델을 선정한다.

전문가 자문을 통해 도출된 방법론 결과에 대한 검토 및 정책제안의 실효성을 검증하고, 이를 통해 다양한 관점에서의 안전사고 예방을 위한 전략을 마련한다.

□ GIS 공간분석 및 통계 분석

건축물의 주소 및 '동'기반 화재 발생현황 데이터의 결합을 통해 사고 발생 요인을 파악하고, GIS 분석도구 및 기계학습을 활용하여 공간데이터 및 공간 네트워크 분석을 실시하고자 한다. 또한 구축된 물리적 환경의 통계분석을 실시하고자 한다.



[그림 1-2] 연구흐름도

출처 : 연구진 작성

3. 선행연구와의 차별성

재난·재해 관련 연구는 재난·재해의 유형, 발생원인, 피해, 대응, 복구, 예방, 예측 등 방대한 분야이다. 본 연구의 목적은 건축물 속성정보와 연계한 화재 및 범죄 데이터를 기계학습을 통해 분석하는 재난재해 위험지역 예측 모델을 구축하는 것이다. 이에 본 연구에서의 선행연구 고찰은 크게 재난·재해 및 화재 취약요소 분석 연구, 도시의 인문·사회·물리적 특성과 화재 발생 연관성 검증 연구, 재난·재해 및 화재의 위험성 평가 연구 분야로 수행하였다.

① 재난·재해 및 화재 취약요소 분석 연구

박무중 외(2011)는 지역별 재난관리를 위한 위험인자(재해위험성, 노출성, 취약성)의 공간분포 해석 연구를 수행하였다. 전국 59개 지점의 35년간 강우량, 기온, 인구자료를 수집하여 경험적 직교함수(Empirical Orthogonal Function)를 활용하였다. 분석 결과 연강수량이나 폭염 등은 인구밀도가 낮은 지역에서 가장 크게 나타나고 있으며, 최대강우량을 보이는 다우 지역은 호우사상의 지속성이 길게 유지되는 특성을 보이고 있다. 또한 확률강우량과 가능최대강수량을 빈도계수로 이용하여 지역별 취약성을 분석하였다.

전상수 외(2008)는 부산시 U-City 모델 구축을 위해 재해 취약지를 분석하였다. 자연재해로 인한 피해평가(행정구역, 시설물, 피해이력)와 사회기반시설 위험도 평가(교량, 터널, 사면, 하천)를 수행한 후 이를 GIS화하여 지역별 취약성을 도출하였다. 장경은 외(2021)는 서울시 중구 화재 취약 지역 분석 연구를 수행하였다. 네트워크 분석 수행을 위해 도로데이터, 소방서 위치, 의료시설 위치, 불법주정차 CCTV 데이터, 노상주차장 데이터를 활용하였으며, 이를 통해 소방서비스 및 의료서비스 권역, 소방차 이동 저해권역 GIS 분석을 통해 화재취약지역을 도출하였다.

진승현 외(2021)는 건축물 용도별 화재위험분석 연구를 수행하였다. 건축물의 규모가 화재 확대에 미치는 영향을 검토하고, 화재 건수 및 연면적을 활용하여 화재위험을 분석하였다. 분석 결과 11층 이상의 건물 및 연면적이 1,000㎡이상 증가했을 때 소규모 화재발생 비율이 증가하는 것을 확인하였다. 권설아 외(2018)는 제천 스포츠센터 화재와 밀양 세종병원 화재참사 사례조사를 통해 물리적·경제적·사회적·환경적 화재 취약성 분석을 수행하였다. 이를 통해 불법건물 증축 및 소방안전시설 점검에 대한 제도의 강화가 필요하다고 하였다.

② 도시의 인문·사회·물리적 특성과 화재 발생 연관성 검증 연구

황종아 외(2020)는 서울시 도시화재 발생지역의 공간분석을 통해 도시쇠퇴 특성과의 관계를 연구하였다. 주거와 비주거지를 구분하여 공간회귀분석을 수행하였으며, 인구·사회적 취약계층이 거주하는 저층 노후 주택지가 화재에 취약한 것으로 나타났다. 이상호(2011)는 도시의 인구사회학적 특성과 화재 발생의 상관성을 연구하였으며, 인구 1만 명당 화재발생 건수에서 고령자 비율, 이혼율, 일인당 차량등록대수가 화재발생과 상관성이 높은 것으로 나타났다.

황성남 외(2011)는 환경위험과 인구적 특성의 관계에 대한 연구를 수행하였다. 홍수, 허리케인, 위험물질 누출에 대한 환경위험지수와 가계의 특성간의 상관관계를 분석하였으며, 연구결과 자연재해는 인구적 특성과의 상관관계가 낮은 반면, 인적재해는 통계적으로 유의미한 상관관계를 나타냈다. 이주호 외(2019)는 도시의 사회, 경제, 환경적 요인에 대한 대형화재 발생 분석 연구를 수행하였다. 시스템 다이내믹스 접근방법을 활용하여 환경적·경제적·사회적 요인인과 화재발생간의 인과지도 모형을 구축하고 이를 통합한 인과지도 모형을 제안하였다.

서민송 외(2020)는 시설물 화재 발생 유의성을 공간계량 분석 방법으로 분석하였다. 진주시를 대상으로 10년간 화재데이터 분석을 위해 선형회귀모형, 공간시차모형, 공간오차모형, LISA 분석을 수행하였으며, 연구 결과 중심상업지역, 공업지역, 주거지역 순으로 화재 발생 빈도가 높은 것을 확인하였다. 화재발생 빈도가 높은 건축물 유형은 제2종 근린생활시설이었으며, 다음으로 단독주택 및 제1종 근린생활시설 순으로 나타났다.

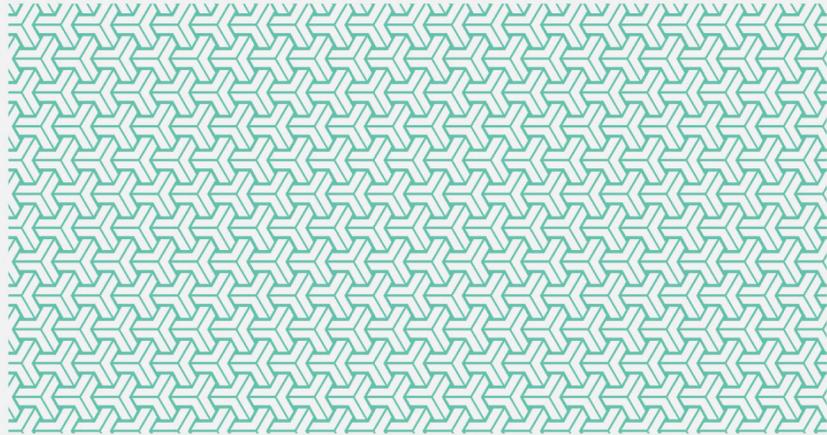
③ 재난·재해 및 화재 위험성 평가 연구

김용문 외(2019)는 QRE평가도구를 활용하여 서울시의 재난 위험도 평가 연구를 수행하였다. 19개 재난(자연재난 3개, 사회재난 16개)을 선정하여 QRE 평가를 수행하였으며, 위험도가 높은 재난은 '자살사고' 및 '대기질 악화'로 나타났고, 위험도가 낮은 재난은 '풍수해', '상수도 누수사고', '수질 오염사고'로 나타났다. 연구자는 QRE 평가도구가 신속한 정책의사결정을 위한 토대를 제공한다고 하였다.

구인혁 외(2016)는 일본의 도시화재 위험성 평가방법의 국내 적용성을 검토하였다. 지구내의 연소성, 소방 활동의 곤란성, 일차 피난 활동의 곤란성, 도로 폐쇄의 가능성, 조시 전체의 연소성, 광역 피난의 곤란성에 대한 평가 항목을 서울시 종로 5가에 적용하였으며, 목조주택 비율이 높은 일본에 비해 철근콘크리트 비율이 높은 한국에의 적용을 위한 건축물 DB의 구축이 필요함을 주장하였다.

이승운 외(2021)는 재해 노출지표와 취약지표를 구성하여 자연재난에 대한 잠재적 재해 위험도 평가체계를 구축하는 연구를 수행하였다. 평가 대상은 사람, 주거·상업시설, 공공시설, 사회복지시설, 농업, 축산업, 도로 등 총 7개로, 이를 국가지점번호 격자단위로 구축하였다. 이 7개 격자단위 구축 결과를 통합하여 서울 수도권 지역과 광역시 등 주요 도시들의 잠재적 위험도가 높게 나타남을 확인하였다.

제2장 재난·재해 빅데이터 관련 동향



1. 국내 재난·재해 빅데이터 시스템
2. 국외 재난·재해 빅데이터 시스템
3. 기계학습 모형 고찰
4. 빅데이터 기반 재난·재해 예측 연구
5. 소결

1. 국내 재난·재해 빅데이터 시스템

1) 국가재난관리정보시스템(NDMS)

「재난 및 안전관리 기본법」에 의한 국가재난관리정보시스템은 2004년 소방방재청에 의해 구축을 추진하였으며, 2015년부터 추진된 통합재난안전정보체계 구축사업을 통해 분산 관리되어 왔던 재난정보를 통합하여 종합적으로 관리하고 있다.

국가재난관리정보시스템은 크게 3개의 포털로 구성되어 있다. 재난관리 업무포털은 중앙정부·지자체·유관기관의 정보를 바탕으로 ‘재난관리 업무지원 시스템’, ‘GIS기반 상황관리시스템’, ‘재난관리자원 공동활용시스템’, ‘소방장비 통합관리시스템’, ‘공공헬기 공동활용시스템’, ‘재난영상 전송시스템’, ‘재난정보 공동이용시스템’으로 구성되어 있다. 모바일 재난관리 포털은 ‘모바일 재난관리 시스템’을 운영하고, 대국민 재난안전포털은 ‘대국민 재난안전 시스템’을 운영하고 있다. 시스템별 소관부서 및 주요기능은 [표 2-1]과 같다.

[표 2-1] 국가재난관리정보시스템 시스템 구성 및 주요기능

구분	시스템(30종)	주요기능	
재난관리업무포털 (22종)	재난(13종)	GIS기반상황관리시스템	자동상황전파, 상황정보 공유, 사고속보관리
		재난관리시스템	시설물 관리, 이재민 관리, 복구계획, 복구진도
		특정관리대상시설시스템	안전점검관리, 시설물 관리
		긴급통신 자원관리시스템	긴급통신자원관리, 공동활용관리
		비상연락망 관리시스템	사회재난 비상연락망, 자연재난 비상연락망
		위기관리매뉴얼시스템	위기관리매뉴얼 관리, 기관 SOP관리
		풍수해보험업무지원시스템	보험관리, 통계관리
		재해구호물자시스템	비축현황관리, 입출고관리, 의연물품관리
		재난정보공동이용시스템	재난정보 송수신 현황관리, 연계 모니터링, 재난정보 이용현황
		재난관리자원공동활용시스템	장비·자재·인력관리, 자원 응원요청관리
		지진재해대응시스템	지진발생정보, 지진피해예측, 지진해일예측, 지진가속도 계측자료 관리
		소하천관리시스템	소하천 정비사업 관리, 소하천 통계
		침수가뭍급경사지시스템	가뭄, 급경사지, 침수흔적정보
	안전(2종)	안전신고정보분석시스템	안전신고 통계관리, 안전신고 상관관계 분석
		국가안전대진단시스템	시설물 관리, 타 시스템 연계

	소방(5종)	소방장비통합관리시스템	소방장비 이력관리, 마감/통계 기능
		소방헬기운항정보시스템	소방헬기 운항정보관리, 실시간 위치추적
		소방항공 무선통신관리시스템	소방상황센터와 항공기 간 무선통신
		경찰헬기 운항정보시스템	경찰헬기 비행계획 및 실시간 모니터링
		재난영상전송시스템	재난영상(소방차량, 모바일) 실시간 모니터링
	해경(2종)	동력수상제거기구관리시스템	동력수상레저 기구관리, 사업관리, 안전검사 관리
		해양교통관리시스템	유도선 지원(선박, 항로, 선착장) 관리
모바일 재난관리 포털(1종)	모바일 재난관리시스템	상황관리, 현장점검, 피해조사, 재해구호, 관측 장비 및 CCTV, 위기관리매뉴얼, 비상연락망, 자원관리	
국민재난안전포털 (6종)	국민재난안전시스템	민방위 대피시설, 재난심리상담, 국민행동요령	
	사유재산피해신고시스템	사유재산피해신고 관리	
	특정관리시설점검시스템	특정관리시설관리, 점검관리	
	복구진도 민간관리시스템	복구진도관리, 사유시설 현황관리	
	재난훈련관리시스템	훈련일정관리, 훈련결과관리	
	방재대책대항자 및 기술인력 관리시스템	사업실적관리, 행정처분 관리	
안전디딤돌(1종)	재난안전정보 포털 앱	재난문자, 국민행동요령, 대피소 정보 등 재난 안전 정보	

출처 : 행정안전부(2018), 2017 재난안전통계연보, 국가재난관리정보시스템 현황, p.180. 직접인용

국가재난관리정보시스템의 시스템별 주요기능을 통해 살펴본 NDMS의 주요 목적은 국가 재난관리에 관한 주요 인력 및 물자의 체계적인 관리와 재난·재해 발생 시 상황을 신속히 전파하고, 대응 및 복구를 체계적으로 수행하는 것이다.

현재까지 국가재난관리정보시스템의 데이터를 활용하여 재난·재해 발생 지역을 예측하는 시스템은 구축되어 있지 않았으며, 행정안전부는 우선적으로 각종 이기종 데이터의 관리 및 연계를 강화하는 방향으로 시스템 개발 사업을 진행하고 있는 것으로 나타났다. 또한(가칭)재난·안전 빅데이터 관리센터 및 분석센터 설립에 대한 논의를 시작하고 있는 단계이기 때문에, 향후 관련 데이터의 체계적인 관리기준이 마련되고 데이터의 공유 및 제공이 원활해지는 시기에, 건축 및 도시공간의 통합안전을 위한 빅데이터 구축 및 연계 방안을 수립하여야 할 것이다.

2) 안전정보통합공개시스템

안전정보통합공개시스템은 재난관리책임기관에서 공개하는 시설 등에 대한 각종 안전점검·진단 등의 결과를 한곳에 모아 통합적으로 공개하는 시스템으로 시설물·건축물의 정보, 점검결과 요약 및 상세제공, 점검결과 이력을 제공하고 있다. 2) 2023년 최종구축을 목표로 하고 있으며 현재는 25개의 주제에 대한 정보를 공개하고 있다.

[표 2-2] 안전정보통합공개시스템 2020년 공개대상 정보

구분	공개정보	담당부처
학교	· 시설명, 위치, 점검자, 점검일 · 종합 안전등급, 지적사항, 시설개선 투자 이력	교육부 교육시설과
승강기	· 승강기번호, 위치, 점검자 · 합격여부, 위반사항, 보완내역	행정안전부 승강기안전과
어린이놀이시설	· 시설번호, 주소 · 안전검사 합격여부, 안전검사일, 유효기간 만료일	행정안전부 안전개선과
체육시설	· 시설명, 위치, 점검자, 점검일 · 점검분야별 안전등급 및 종합 안전등급, 중대 결함 내역 및 총평, 점검분야별 소견, 조치사항	문화체육관광부 스포츠산업과
전기	· 시설명, 점검종류, 점검일 · 적합여부	산업통상자원부 에너지안전과
어린이집	· 시설명, 위치, 점검일 · 전기안전점검 적합 여부, 보험가입 현황, 환경·위생(식중독 발생, 실내 공기질 등) 현황	보건복지부 보육기반과
국립공원	· 시설명, 상태, 종합 안전등급, 조치상태 및 계획	환경부 자연공원과
유해화학물질	· 시설명, 취급자 인적사항, 취급현황 · 법령위반사항, 사고발생 사례	환경부 화학안전과
건축물 기본정보	· 대지면적, 연면적, 건축면적, 건폐율, 용적률, 허가일, 사용승인일	국토교통부 건축정책과
건축물 안전정보	· 시설명, 위치, 점검자, 점검일 · 점검분야별 점수, 보수·보강 조치 결과, 점검결과에 대한 평가, 화재안전 성능보강 결과	국토교통부 건축정책과
여객선	· 선명, 선종, 선령, 정원, 톤수 등 선박명세 및 검사일, 검사자 · 합격여부, 안전규정 위반이력, 해양사고이력, 운항관리규정 심사일 및 세부사항	해양수산부 연안해운과
항만	· 시설명, 위치, 점검자, 점검일 · 종합 안전등급	해양수산부 항만기술안전과

2) 안전정보 통합공개시스템 홈페이지, https://safewatch.safemap.go.kr/main/contentsView.do?cntntsId=CONTENTS_00000 (검색일 : 2022.06.20.)

전통시장	· 시설명, 위치, 점검자, 점검일 · 주요 지적사항, 종합 안전등급, 점검조건, 시장별 점검결과 보고서	중소기업벤처부 시장상권과
사방	· 관리번호, 위치, 점검자, 점검일 · 종합의견, 보수 필요사항, 조치결과, 조치내용, 사업기간	산림청 산사태방지와
시설물안전	· 시설명, 위치, 점검자 · 시설점검결과, 시설제원, 보수보강결과, 중대결함 관리내용	국토교통부 건축정책과
액화가스	· 시설명, 위치 · 검사종류명, 검사일자, 검사결과	산업통상자원부 에너지안전과
유원	· 시설명, 위치 · 유원시설업 종류, 영업소상태, 검사종류, 검사일자, 검사결과	문화체육관광부 관광산업정책과
병원	· 시설명, 위치 · 의료기관명, 인증 등급, 인증 유효기간	보건복지부 의료기관정책과
식품판매	· 시설명, 위치 · 위반일자, 위반내용, 처분내용	식품의약품안전처 소비자위해예방정책과
건설안전	· 공사기간, 사고원인, 신구사유, 사례, 물적피해 · 인적피해, 단계별저감대책	국토교통부 건설안전과
원자력 발전	· 정기검사 호기, 차수, 기간, 사고현황	원자력안전위원회 안전정책과
소규모 공공	· 시설명, 시설구분 · 공공시설여부, 위험시설지정여부, 점검일자, 처리 결과	행정안전부 재난경감과
폐기물처리	· 시설명, 위치, 승인일자 · 지하수오염여부 확인방법, 검사횟수, 정기검사, 유지보수횟수	환경부 폐자원에너지과
농업생산기반	· 시설명, 위치 · 농업생산시설별 제원, 점검일자, 점검결과	농림축산식품부 농업기반과
해체공사장	· 시설명, 위치, 구조, 주용도 · 공사기간, 감리자, 시공자	국토교통부 건설정책과

출처 : 안전정보 통합공개시스템 홈페이지, https://safewatch.safemap.go.kr/main/contentsView.do?cntntsId=CONTENTS_00000 (검색일 : 2022.06.20.)

위와 같이 제공된 정보는 안전정보 통합공개시스템 홈페이지 내 안전점검정보 지도와 생활안전지도로 분리되어 서비스 되고 있다. 안전점검정보지도는 교육/보육, 보건/의료, 숙박시설, 주거시설, 상업/판매, 문화시설, 체육시설, 식품위생, 산업시설, 공공시설, 기반시설에 대해 건축물 현황(층수, 용도, 구조, 연면적, 사용승인일) 및 안전점검결과(시설점검, 전기점검, 승강기점검, 식품위생위반사항, 안전대진단, 인증결과, 실내공기질 등)를 세부적으로 공개하고 있다.



[그림 2-1] 건축물 및 시설물 안전점검정보 지도 예시

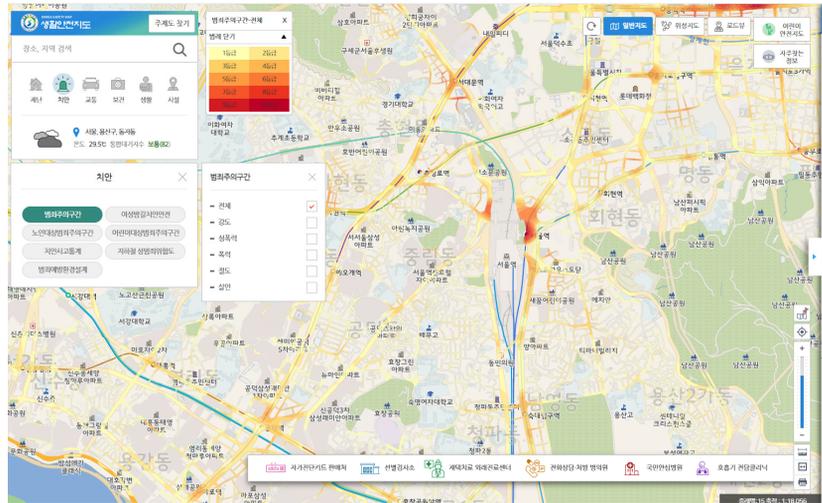
출처 : 안전정보통합공개시스템 홈페이지, <https://safewatch.safemap.go.kr/safty/map.do#>
(검색일 : 2022.06.20.)

생활안전지도에서는 6개의 안전정보통합관리 분야(재난, 치안, 교통, 보건, 생활, 시설)에 대한 안전정보를 제공하고 있다. 안전점검정보지도와는 달리 민감정보에 대한 것은 분포도로 보여주고 있으며, 안전에 필요한 정보(병원 등)에 대한 정보는 지도에 상세히 제공하고 있다.

[표 2-3] 생활안전지도 주제별 공개정보 내역

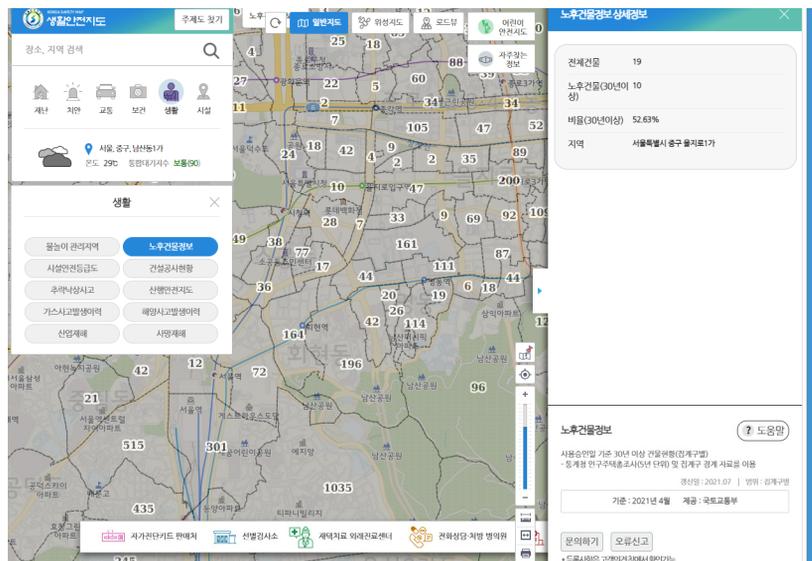
구분	공개정보
재난	실시간 지진발생정보, 지진발생이력, 붕괴발생이력, 산불발생이력, 산사태발생이력, 산사태위험도, 송수범람위험도, 해안침수예상도, 침수흔적도, 열분포도, 비행금지구역, 재난배상책임보험
치안	범죄주의구간(전체, 강도, 성폭력, 폭력, 절도, 살인), 여성범죄치안안전(전체, 강도, 절도 폭력, 성폭력), 노인대상 범죄주의구간, 어린이대상 범죄주의구간, 치안사고통계, 지하철 성범죄위험도, 범죄예방방화경설계
교통	실시간 교통정보, 사고다발지역, 교통사고정보(클러스터 분석, 밀도분석, 시계열 서비스), 대형교통사고 발생현황, 도로시설, 졸음쉼터, 상습결빙구간, 자전거길
보건	실시간 보건정보(통합대기질수, 미세먼지, 초미세먼지, 이산화질소, 오존, 자외선지수, 식중독지수, 동파가능지수), 질병예측정보(감기, 눈병, 식중독, 천식, 피부염), 병의원, 약국, 응급의료시설, 주의보발령이력, 어린이 천식, 어린이 아토피, 식중독 발생통계(원인 및 시설), 가축전염병, 철새도래지, 자살발생통계
생활	물놀이 관리지역, 노후건물정보, 시설안전등급도, 건설공사현황, 추락낙상사고, 산행안전지도, 가스사고 발생이력, 해양사고 발생이력, 산업재해, 사망재해
시설	실시간 시설정보, 공공(관공서, 교육환경보호구역, 소방시설, 유아시설, 치안시설, 학교), 교통(가로등, 방범등, 전기충전소, 주유시설), 재난안전(자동심장충격기, 무더위쉼터, 안전비상벨, 제설함, 해안방파시설), 대피시설, 복지시설, 유해화학시설, 이동통제 초소, 기타(승강기 시설정보, 어린이놀이시설 사고정보, 우수다중시설, 편의점)

출처 : 안전정보 통합공개시스템 생활안전지도, <https://www.safemap.go.kr/main/smmap.do?flag=2> (검색일 : 2022.06.20.)



[그림 2-2] 안전정보통합공개시스템의 생활안전지도 범의주요구간 예시

출처 : 안전정보통합공개시스템 생활안전지도, <https://www.safemap.go.kr/main/smap.do?flag=2>(검색일 : 2022.06.20.)



[그림 2-3] 안전정보통합공개시스템의 생활안전지도 노후건물정보 예시

출처 : 안전정보통합공개시스템 생활안전지도, <https://www.safemap.go.kr/main/smap.do?flag=2>(검색일 : 2022.06.20.)

3) 국가화재정보시스템(NFDS)

국가화재정보시스템은 화재의 원인·발화·발견·통보 및 연소확대 등의 화재발생부터 피난상황, 소방설비의 작동 등 화재진압까지의 화재 매커니즘과 관련한 화재정보와 이를 통계화한 정보를 국민에게 제공하고, 소방청을 중심으로 유관 기관 및 산·학·연 전문기관 간에 지식 공유 및 정보의 장을 제공하기 위한 목적으로 구축되었다.³⁾

구축 목적은 ① 시스템을 통해서 소방관계자뿐만 아니라 개개인이 예보된 기상 조건 및 화재발생인자 분석정보에 따라 지역별 화재발생위험도를 확인하고 예방할 수 있는 프로그램 구축하고, ② 화재발생 주요인자 감지 및 지역별 화재발생 위험도 정보를 실시간 제공하며, ③ 실시간 이상징후 감지 및 모니터링 기능을 통해 국민 및 유관기관에게 화재위험 경보를 함으로써 각 기관별 화재예방정책 수립하고, ④ 화재발생 위험성을 사전에 인지하고 화재예방 및 대응자료로 활용하는 것이다.⁴⁾

주요 제공 정보는 실시간 화재발생현황, 화재현황 분석, 화재통계, 화재관련 논문·사례·보고서 등이다. 특히 화재에 대한 상세한 통계를 제공하고 있으며, 각 통계항목별로 인명피해(사망, 부상) 및 재산피해(부동산, 동산) 통계를 제공하고 있어 건축물의 경우 어떠한 구조에서 어떠한 유형의 화재가 발생하는지 파악하기 용이하다.

[표 2-4] 화재통계 제공 내역

구분	통계제공 목록
화재유형	건축·구조물, 자동차·철도차량, 위험물·가스제조소 등, 선박·항공기, 임야, 기타 (쓰레기 화재 등)
발화열원	작동기기, 담뱃불·라이터불, 마찰·전도·복사, 불꽃·불티, 폭발물·폭파, 화학적 발화열, 자연적 발화열, 미상, 기타
발화요인	전기적 요인, 기계적 요인, 화학적 요인, 가스누출(폭발), 교통사고, 부주의, 자연적 인 요인, 방화, 방화의심, 제품결함, 미상, 기타
발화기기	계절용 기기, 생활기기, 주방기기, 영상·음향기기, 사무기기, 조명·간판, 배선·배선기구, 전기설비, 산업장비, 농업용장비, 의료장비, 상업장비, 차량·선박부품, 드론, 미상, 기타
최초착화물	가구, 침구·직물류, 종이·목재·건축 등, 합성수지, 간판·차양막 등, 식품, 전기·전자, 위험물 등, 가연성 가스, 자동차·철도차량·선박·항공기, 쓰레기류, 미상, 기타
건물구조	스트레가, 한식기와, 컨테이너, 블록조, 슬라브가, 초가, 샌드위치패널, 비닐하우스, 석조, 와가, 시멘트 기와, 칼라피복철판, 철근콘크리트조, 기타

출처 : 국가화재정보시스템 홈페이지, <https://nfds.go.kr/stat/general.do> (검색일 : 2022.06.20.)

3) 국가화재정보시스템, <https://nfds.go.kr/intro/introduction.do> (검색일 : 2022.06.20.)

4) 국가화재정보시스템, <https://nfds.go.kr/intro/introduction.do> (검색일 : 2022.06.20.)

국가화재정보시스템에서 제공하는 화재통계는 크게 일반유형, 발화장소 및 지점과 지역별로 구분된다. 일반유형은 화재유형, 발화열원, 발화요인, 발화기기, 최초착화물, 건물구조별 화재통계를 제공하고, 발화장소 및 지점은 발화장소, 발화지점, 차량화재장소, 차량발화지점, 임야화재장소, 임야발화지점별 화재통계를 제공하고 있으며, 지역은 행정구역별 및 소방관할관서별 화재통계를 제공하고 있다. 화재통계 세부분류는 발화요인 및 화재장소별 세부분류 기준으로 통계를 구축하고 있다.

특히 건축물에서의 화재장소는 동종·유사 성격을 건축물 허가상의 주 용도시설군으로 분류한 대·중분류로 구분하고, 소방청에서 자체적으로 시설내의 기능장소 용도로 분류한 소분류로 분류한다.⁵⁾ 이에 국가화재정보시스템에서 국민에게 제공하는 화재데이터 중 건축물 용도별 화재발생 데이터와 구조별 화재발생 데이터를 결합하면 건축물의 용도별·구조별로 화재위험 건축물을 도출할 수 있을 것이다.

[표 2-5] 화재 발화요인별 세부요인

발화요인	세부요인		
전기적 요인	<ul style="list-style-type: none"> • 누전 및 지락 • 접촉불량에 의한 단락 • 절연열화에 의한 단락 	<ul style="list-style-type: none"> • 과부하/과전류 • 미확인 단락 • 반단선 	<ul style="list-style-type: none"> • 압착 및 손상에 의한 단락 • 충전단락 • 트래킹에 의한 단락
기계적 요인	<ul style="list-style-type: none"> • 과열, 과부하 • 오일 및 연료누설 • 자동제어 실패 	<ul style="list-style-type: none"> • 수동제어 실패 • 정비불량 	<ul style="list-style-type: none"> • 노후 • 역화
제품결함	<ul style="list-style-type: none"> • 설계상 결함 	<ul style="list-style-type: none"> • 제조상 결함 	
가스누출	<ul style="list-style-type: none"> • 가스누출(폭발) 		
화학적 요인	<ul style="list-style-type: none"> • 화학적 폭발 • 금속성물질의 물과 접촉 	<ul style="list-style-type: none"> • 화학적 발화(유증기 확산) • 자연발화 	<ul style="list-style-type: none"> • 혼촉발화
교통사고	<ul style="list-style-type: none"> • 교통사고 		
부주의	<ul style="list-style-type: none"> • 담배꽂초 • 음식물 조리중 • 불장난 • 용접, 절단, 연마 	<ul style="list-style-type: none"> • 불씨, 불꽃, 화원방치 • 쓰레기 소각 • 빨래삶기 • 가연물 근접방치 	<ul style="list-style-type: none"> • 논, 임야 태우기 • 유류취급 중 • 폭죽놀이
자연적인 요인	<ul style="list-style-type: none"> • 자연적 재해 	<ul style="list-style-type: none"> • 돌보기 효과 	
방화	<ul style="list-style-type: none"> • 방화 	<ul style="list-style-type: none"> • 방화의심 	

출처 : 소방청(2021), 2020년도 화재통계연감, p.23. 연구진 재구성

5) 소방청(2021), 2020년도 화재통계연감, p.33.

[표 2-6] 화재통계에서의 화재장소 분류체계

대분류	중분류
주거시설	단독주택, 공동주택, 기타주택
교육시설	학교, 연구, 학원
판매·업무시설	판매시설, 공공기관, 일반업무, 숙박, 청소년시설, 군사시설, 교정시설
집합시설	관람장, 공연장, 종교시설, 전시장, 운동시설
의료복지시설	건강시설, 의료시설, 노유자시설
산업시설	공장시설, 창고, 작업장, 발전시설, 지중시설, 동식물시설, 위생시설
운수, 자동차, 문화재시설	자동차시설, 향만시설, 역사, 터미널, 문화재
생활서비스	위락시설, 오락시설, 음식점, 일반서비스
자동차, 철도	자동차, 농업기계, 건설차량, 군용차량, 철도차량
위험물 가스 제조소 등	위험물 제조소 등, 가스제조소 등
선박및항공기등	선박, 항공기
임야 및 기타	산불, 들불, 야외, 도로, 기타건축물(공중화장실 등)

출처 : 소방청(2021), 2020년도 화재통계연감, pp.33~36. 연구진 재구성

4) 경찰청 지리적 프로파일링 시스템(GeoPros)⁶⁾

경찰청은 2009년 4월부터 기존의 범죄정보관리시스템(CIMS), 수사종합검색시스템(CRIFISS), 과학적 범죄분석시스템(SCAS) 등을 통해 축적된 범죄 자료를 기반으로 공간통계기법을 적용한 지리적 프로파일링 시스템을 구축하여 운영하고 있다(이문국 2015, p.74).

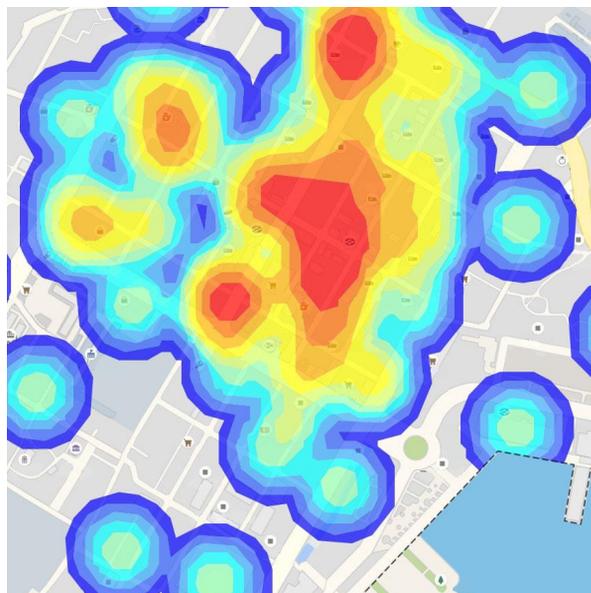
초기 GeoPros 시스템은 개발 당시 현재 Hot Spot라고 알려져 있는 범죄다발지역의 분석, 수사대상자 및 사건에 대한 분석, 연쇄 범죄자의 거주지 예측 등 범죄 및 범죄자의 지리적 특성과 관련된 다양한 기능을 갖추고 있었다(경찰백서 2010, p.181). 특히 Hot Spot은 범죄발생의 밀집도 분석을 시각화 하여 시각화 하여 범죄 유형 및 시간대별 맞춤형 방법 전략 수립에 도움이 되었다(경찰백서 2010, p.182).

이후 2013년부터 GeoPros 고도화 사업을 시행하여 범죄 자료뿐만 아니라 지역 특성, 주거 형태, 연령대, 소득 수준 및 유동 인구 등의 인구사회학적 특성을 분석 변인에 추가하였다. 이를 통해 범죄예측력을 높였다.

6) 박유나 외(2020), 한국형 범죄예방환경설계를 위한 장소프로파일링 기법 개발 연구, pp.28-29.

GeoPros의 분석은 크게 ‘범죄환경진단’, ‘범죄다발분석’, ‘사건분석’, ‘수사대상 분석’, ‘지하철디지털성범죄’, ‘현황분석’으로 구분되어 진다. 범죄환경진단은 해당 지역이 범죄 발생에 있어 얼마나 취약한지를 단계로 구분하여 시각화로 보여주는 기능으로, 범죄위험도와 관서별 범죄분포, 사용자 지도분석으로 구성된다. 범죄다발분석은 기존의 HotSpot 분석과 더불어 112신고분석의 기능을 포함하고 있으며, 사건분석은 형사사법정보시스템(KICS) 데이터를 활용한 특정사건의 상세분석 기능을 제공한다. 수사대상분석은 사건 발생 시 우선적 수사대상자를 선별하기 위한 것으로 일부 사용자에게 제한적으로 사용되고 있다. 지하철디지털성범죄 분석은 KT의 유동인구 데이터와 KICS 성범죄 관련 사건 데이터를 기반으로 지하철역 및 주변 성범죄 위험도를 제공한다.

주로 사용되는 범죄환경진단의 범죄위험도 측정은 5대 강력범죄 중 살인을 제외한 4가지 죄종의 범죄 유관요인과 이를 바탕으로 추정된 범죄발생에 대한 위험도를 국가기초구역 단위로 표시하고 있다. 분석에 사용되는 데이터는 경찰청 KICS의 범죄발생 건수, 인구주택총조사의 주거 인구 수, 유동인구, 유흥업소, 전과자, 가로등, CCTV, 편의점 수 및 위치정보이다. 지방경찰청 및 관할 경찰서는 해당 범죄위험도 분석 자료를 바탕으로 범죄발생지역 및 시간대를 예측하여 순찰 강화를 통해 범죄를 사전에 예방하고 있다. 또한 사건 발생 시 범인의 도주로 예측, 최단거리 범행 장소 도착, 범죄 통계 분석 등을 수행하고 있다.



[그림 2-4] GeoPros 범죄위험지역(Hot Spot) 분석 예시

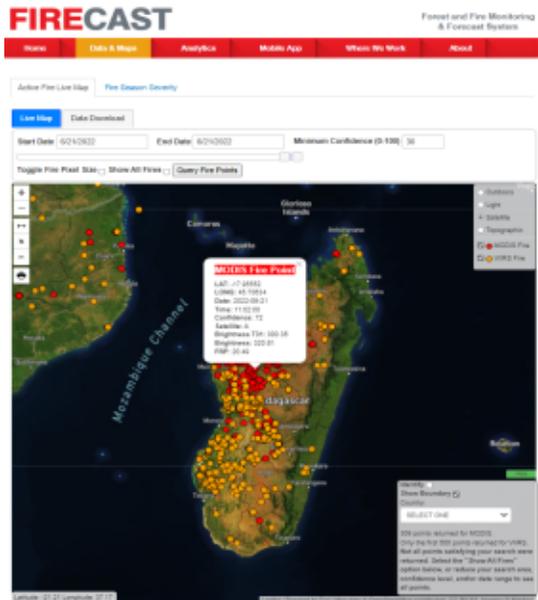
출처 : 경찰청 비공개 내부자료

2. 국외 재난·재해 빅데이터 시스템

1) FIRECAST by Conservation International⁷⁾

FIRECAST는 산림·화재 모니터링 및 예측 시스템으로 사용자의 특정 요구에 대응하는 NRT(near real-time) 모니터링을 제공하는 자동 분석 및 경고 시스템이다. CI(Conservation International)에 의해 볼리비아, 콜롬비아, 에콰도르, 인도네시아, 마다가스카, 페루, 수리남 총 7개 국가에서 운영되고 있다.

FIRECAST는 위성 관측을 통해 화재, 화재위험환경, 산림벌채 등 생태계 위협 요인들을 추적한 긴급정보를 이메일 알림, 지도, 보고서 등으로 고객에게 제공한다. 메일에는 화재 위치 또는 화재 위험 지도와 Google 어스로 데이터를 가져오기 위한 KML(지리 데이터를 표시 파일)을 포함한다.



[그림 2-5] Active Fire Live Map

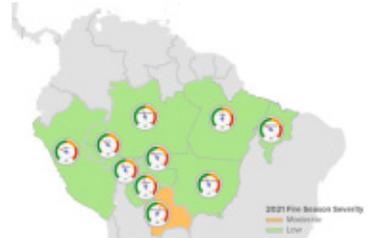
출처 :

<https://firecast.conservation.org/DataMaps/LiveView> (검색일 : 2022.06.22.)

FIRECAST 시스템의 목표는 화재 관리 및 예방, 산림감시 및 모니터링, 보전지역관리, 지속가능한 토지이용계획에 이르기까지 현장 중심의 의사결정을 지원하고, 실시간 NRT를 통해 전략적 토지 이용 및 보존 정책을 지원하기 위해 생태계 동향에 대한 유용한 정보를 제공하는 것이 목표이다. FIRECAST는 지방 및 중앙정보, 국가보존기구, 지역사회, 펀딩 에이전시, 보존지역 매니저들을 대상으로 설계되었으며, FIRECAST에서 제공하는 주요 서비스는 아래와 같다.

7) <https://firecast.conservation.org/>을 참고하여 재작성

[표 2-7] FIRECAST의 주요 서비스

주요 서비스	내용
<p>Active Fire Detection</p> 	<ul style="list-style-type: none"> • Firecast는 NASA의 아쿠아 및 테라 위성에 탑재된 MODIS(Moderate Resolution Imaging Spectroradiometer)와 VIIRS (Visible Infrared Imager Radiometer Suite) 센서를 통해 전송되는 데이터를 통해 화재 감지 정보를 실시간으로 제공
<p>Fire Season Severity Forecasting</p> 	<ul style="list-style-type: none"> • 화재 기간 심각도 예측은 대서양과 태평양의 해수면 온도와 남미 전역의 화재 발생률을 분석하여 예측 • 잠재적인 화재 기간 심각도에 대해 미리 인지함으로써 화재관리 및 예방에 유용하게 활용 가능

출처 : <https://firecast.conservation.org/About/DataDescription> (검색일 : 2022.06.22.)
<https://firecast.conservation.org/DataMaps/FireSeasonSeverity>(검색일 : 2022.06.22.)

2) Fire Danger Forecast by USGS(U.S. Department of the Interior)⁸⁾

Fire Danger Forecast는 USGS와 미국 산림청이 함께 운영하는 화재위험예측 프로젝트로 미국 내 잠재적 화재 위험성을 모니터링하여 디지털 지도로 개발하는데 중점을 두고 있다. 화재와 화재 확산 가능성을 예측하는 디지털 지도 시리즈를 자동적으로 생성하는 프로세스를 유지 및 운영한다.

Fire Danger Forecast는 위성자료 기반의 식생 지수, 다양한 생물지리학적 변수 및 기상 정보를 조합하여 화재 가능성 지수, 대형 화재 확률, 화재 확산 확률에 대한 7일 동안의 예측 지도와 데이터를 제공한다. 이 자료들은 뷰어 어플리케이션, 지도 및 데이터 제공 페이지, GACC 지역 예측 그래프 페이지, 데이터 서비스 페이지를 통해 시각화 및 다운로드가 가능하며, 매일 업데이트되어 제공되고 있다.

8) <https://www.usgs.gov/fire-danger-forecast/about-fire-danger-forecast>을 참고하여 재작성

Fire Danger Map and Data Products

The Map and Data Products page offers bulk download of the Fire Danger Forecast data suite.

Wildland Fire Potential Index Maps

WFPI Forecast - Day 1

WFPI Forecast - Day 2

WFPI Forecast - Day 3

WFPI Forecast - Day 4

Daily Period

Year: 2022

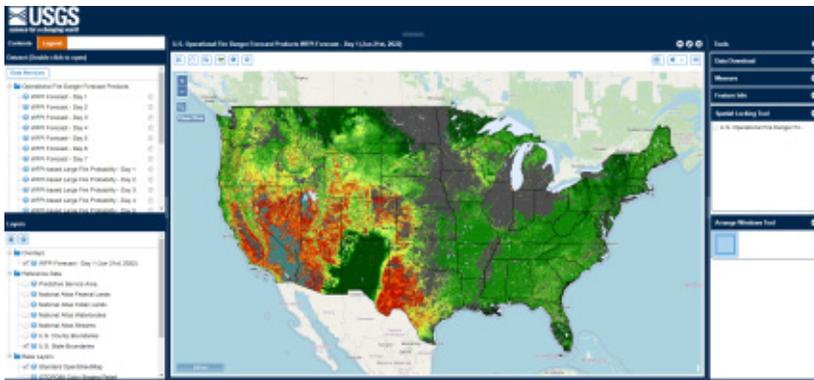
Month: Jun

Day: 01

[Bulk Download](#)

[그림 2-6] Fire Danger Forecast에서 제공하는 최근 7일 동안의 WFPI 지도

출처 : <https://firedanger.cr.usgs.gov/apps/staticmaps> (검색일 : 2022.06.22.)

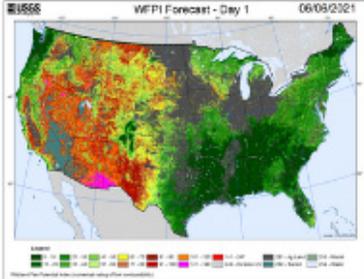
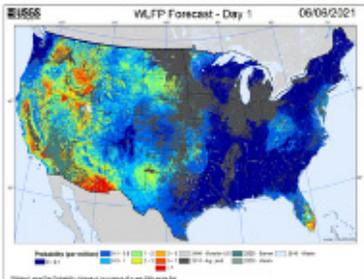
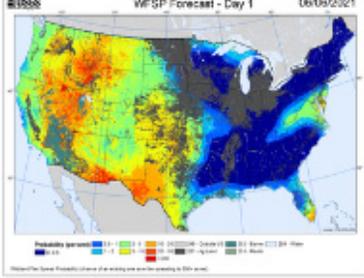
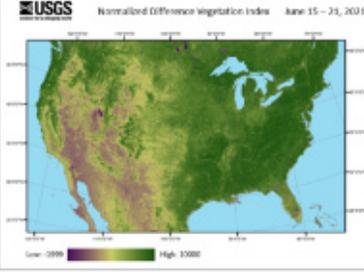


[그림 2-7] Fire Danger Viewer Application

출처 : <https://firedanger.cr.usgs.gov/viewer/index.html> (검색일 : 2022.06.22.)

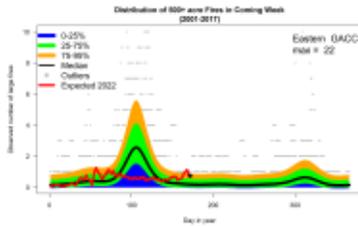
Fire Danger Forecast에서 제공하는 주요 서비스는 Wildland Fire Potential Index (WFPI), WFPI-based Large Fire Probability (WLFP), WFPI-based Fire Spread Probability (WFSP), Fire Danger Base Vegetation Indices, Fire Danger GACC Regional Forecast Graphs 이다.

[표 2-8] Fire Danger Forecast의 주요 서비스

주요 서비스	내용
<p>Wildland Fire Potential Index (WFPI)</p> 	<ul style="list-style-type: none"> WFPI는 수분 및 바람 기반 식생 가연성 지표 현재 식생의 녹색도(녹지, 현재 10-h dead fuel moisture, 풍속, 비, 건구 온도의 비율)를 표시 식생이 대부분 또는 완전히 건조된 상태, 10-h dead fuel moisture가 낮고, 풍속이 높을 때 가연성이 높음 반대로 식물이 완전히 녹색에 가깝고 dead fuel moisture가 높고, 풍속이 낮을 때 가연성이 가장 낮음
<p>WFPI-based Large Fire Probability (WLFP)</p> 	<ul style="list-style-type: none"> WLFP는 특정 시간 및 위치에서 대형 화재(500에이커 이상)가 발생할 확률을 표시 이 지도에서 사용된 통계 모델은 WFPI, 가용 가능한 연소물질의 위치, 연소물질 유형 및 화재의 계절적 패턴 날짜를 기반으로 생성
<p>WFPI-based Fire Spread Probability (WFSP)</p> 	<ul style="list-style-type: none"> WFSP는 1 에이커의 화재가 500에이커 이상을 태우는 화재로 확산될 가능성을 표시 화재 위험의 계절적 및 위치적 차이는 주로 WFPI에 사용된 변수를 통해 계산
<p>Fire Danger Base Vegetation Indices</p> 	<ul style="list-style-type: none"> 화재 위험 예측 식생 지수는 매주 업데이트되는 NDVI(Normalized Difference Vegetation Index)와 RG(Relative Greenness)에 기반 NDVI는 특정 픽셀의 근적외선 신호(식물이 반사)와 적색 신호(식물이 흡수) 간의 차이를 측정 RG는 과거 NDVI의 관찰 범위의 픽셀과 비교하여 동일 픽셀의 녹색 정도를 표시

Fire Danger GACC Regional Forecast

Graphs



- 미국을 9개 지역으로 구분하여 지역별로 예전 정보와 비교하여 가장 최근의 화재 위험에 대한 전망을 보여주는 예측 그래프
- X축은 1년을 일단위로 표시, Y축은 대형 화재의 수를 나타냄

출처 : <https://www.usgs.gov/fire-danger-forecast/about-fire-danger-forecast>(검색일 : 2022.06.22.)
<https://firedanger.cr.usgs.gov/apps/gacc> (검색일 : 2022.06.22.)

3) 건축물 화재 위험도 예측 모델(Metro21: Smart Cities Institute)⁹⁾

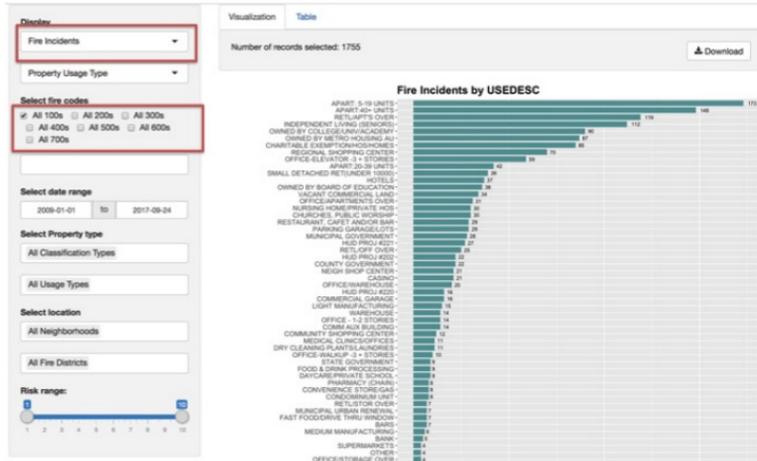
Metro21: Smart Cities Institute는 카네기 멜론 대학에서 수행하는 프로젝트로 최신 기술연구와 개발을 통해 5개 분야(Smart Infrastructure, Safe Mobility, Net Zero Energy, Equitable Economic Prosperity, Clean Air & Water)에서 도시 문제들을 해결하기 위한 여러 프로젝트를 진행하고 있다. 이 중 건축물 화재 위험도 예측 모델은 특정 건축물을 대상으로 6개월 내 화재 발생 가능성을 예측하는 프로젝트이다. 화재 위험도를 예측하기 위하여 2009년에서 2017년까지 미국 피츠버그에서 발생한 화재 데이터, 건축물 가치 측정 자료, 건축물 유형(규모, 가격), 토지 정보, 그 외 소음 및 시설 위반사항 데이터를 이용하여 모델을 구축하였다.

예측 모델은 20,806동의 비주거 건축물(상업, 공공, 제조업 건축물 등)에 대해 1부터 10까지 화재 위험도를 측정하고, 다양한 용도의 건축물 중 고위험 건축물을 “Burgh’s Eye View”¹⁰⁾을 통해 공간정보로 시각화하여 제공하였다. 예측 결과, 167동의 건축물이 고위험도(7점~10점), 596동이 중위험도(4점~6점), 나머지 건축물은 모두 저위험도(1점~3점)로 나타났다. 건축물 유형으로 살펴보면 40 유닛 이상의 아파트, 5~19유닛의 아파트, 교육 위원회 소유 건축물, 자선단체 소유 건축물, 고령자 거주 건축물이 위험한 것으로 나타났다.

9) <https://www.cmu.edu/metro21/projects/smart-infrastructure/fire-risk-analysis.html>; Madaio, Michael A(2018), “Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections,”을 참고하여 재작성

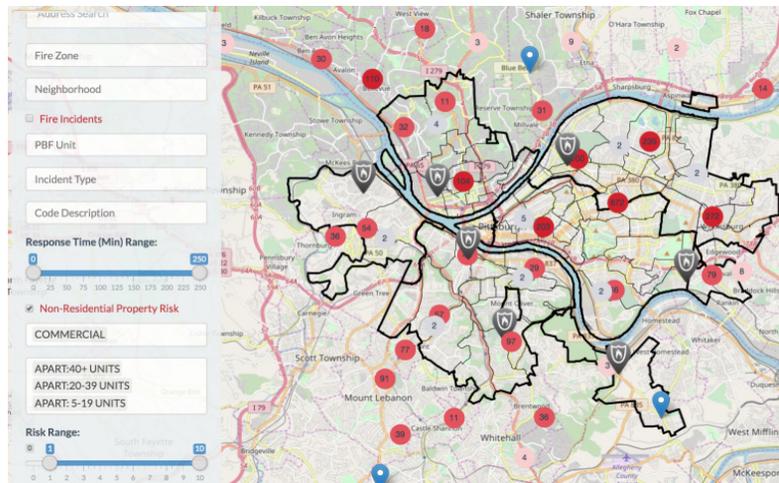
10) 피츠버그시에서 운영하는 공간정보시각화 툴 (<http://www.wprdc.org/showcase/burghs-eye-view/>)

모델이 구축된 후 예측 정확성에 대해 살펴본 결과 건축물 화재 사건 45건(31%) 중 14건이 중간 이상의 화재 위험도로 예측된 건축물에서 발생했으며, 이는 평균 화재사고(0.20%)보다 높은 것으로 나타났다. 이 모델을 통해 생성되는 최첨단 예측 정보는 매주 업데이트되어 소방국에 제공되며, 데이터 대시보드 및 인터랙티브 맵에 예측 정보를 나타냄으로써 지역, 화재구역, 건축물 유형 간에 화재 위험도를 쉽게 비교하고 화재 위험에 대비가 가능하다.



[그림 2-8] 건축물 유형에 따른 예측 정보를 보여주는 데이터 대시보드

출처 : Madaio, Michael A(2018), "Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections, p.25



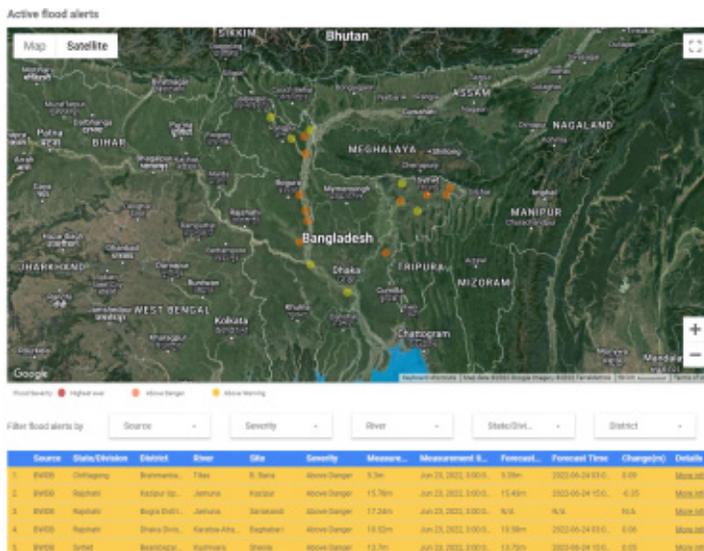
[그림 2-9] 고위험 건축물을 보여주는 인터랙티브 맵

출처 : Madaio, Michael A(2018), "Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections, p.25

4) 구글(Google) 홍수예측 시스템¹¹⁾

구글의 홍수예측 시스템은 홍수위험도가 높은 지역의 거주민들에게 홍수 정보를 제공하는 서비스로 실시간 데이터를 기반으로 수시 업데이트되는 최신 정보를 제공한다. 2018년 인도 파트나 지역을 대상으로 서비스가 제공되었고 현재는 대상 지역이 확대되어 인도 전역과 방글라데시 일부 지역이 해당한다.

홍수예측 시스템은 실시간 강 수위 예측, 강수량 예측, 현지 데이터를 지속적으로 모니터링하며, 이 모니터링 정보들을 기반으로 2개의 모델을 통해 홍수를 예측한다. 첫 번째는 수문학적 모델로 미래의 강 수위를 예측하며, 두 번째 모델인 침수 모델은 홍수 지역을 예측하여 구글맵에서 홍수예측 지도를 제공한다. 또한 구글은 홍수 예측의 정확성을 향상시키기 위해 HydroNets이라는 차세대 홍수예측 시스템을 개발하였으며, HydroNets은 홍수경보 리드타임을 두 배로 늘려 발생 시간의 90%, 강 수위 15cm 오차범위 내에서 홍수를 예측한다.¹²⁾ 이 정보를 기반으로 스마트폰 위치 서비스가 켜져있고 홍수 발생 또는 예상지역에 있는 사람들에게 홍수예보알림이 간다.

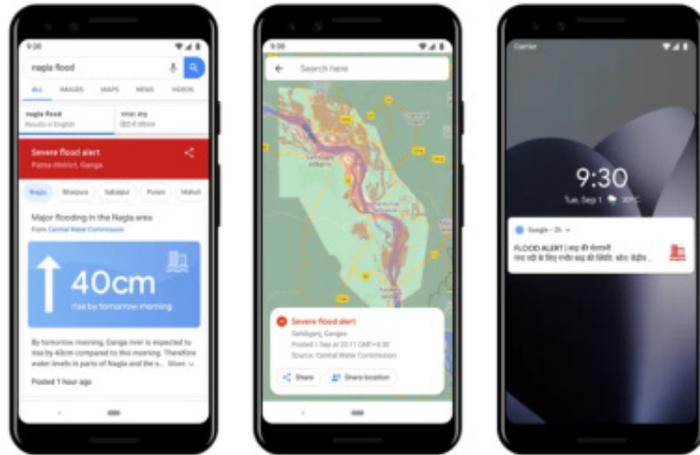


[그림 2-10] 구글 홍수 알림 대시보드

출처 : <https://datastudio.google.com/reporting/9526ada3-68c8-4a5b-8661-4b1f61e6a6d5/page/CpieB?s=mhiVWV6FPes>(검색일 : 2022.06.23.)

11) https://support.google.com/publicalerts/topic/11030989?hl=en&ref_topic=2696427를 참고하여 재작성
 12) <https://ai.googleblog.com/2020/09/the-technology-behind-our-recent.html> (검색일 : 2022.06.23.)

예일대의 Economic Growth Center와 구글이 협력하여 갠지스-브라마푸트라 (Ganges-Brahmaputra)를 대상으로 수행한 연구에 따르면 홍수예보알림을 받은 70%의 가구는 홍수가 발생하기 전에 알림을 받았으며, 알림을 받은 가구 중 65%는 홍수에 대비한 것으로 나타났다.¹³⁾



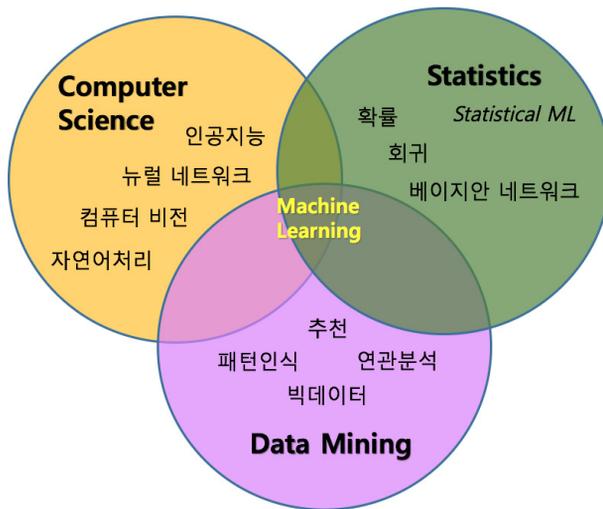
[그림 2-11] 구글 홍수 예측 모바일 앱

출처 : <https://www.theverge.com/2020/9/1/21410252/google-ai-flood-warnings-india-bangladesh-coverage-prediction>(검색일 : 2022.06.23.)

13) <https://egc.yale.edu/pande-and-coauthors-using-technology-save-lives-during-indias-monsoon-season>(검색일 : 2022.06.23.)

3. 기계학습 모형 고찰

기계학습은 컴퓨터를 인간과 같이 학습시켜 새로운 데이터에 대한 대응이 가능하도록 만드는 일련의 알고리즘이나 기술을 의미한다. 즉, 기계학습은 컴퓨터가 사전에 프로그래밍된 방법으로 문제에 대응하지 않고, 데이터로부터 대응 방법을 학습하여 문제에 대응할 수 있다. 컴퓨터의 학습과정을 위해 데이터가 필요하며, 양질의 데이터 확보는 기계학습 모형의 성능을 향상시킬 수 있다. 기계학습은 일반적으로 인공지능과 자연어 처리와 같은 컴퓨터 사이언스, 회귀분석 및 확률계산과 같은 통계, 패턴인식과 빅데이터 분석 등 데이터마이닝 분야에 주로 사용된다.



[그림 2-12] 머신러닝의 주요 활용분야

출처 : 텐서플로우 블로그, https://tensorflow.blog/%ED%95%B4%EC%BB%A4%EC%97%90%EA%B2%8C-%EC%A0%84%ED%95%B4%EB%93%A4%EC%9D%80-%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-1/hackers_resource/ (검색일 : 2022. 06.15.)

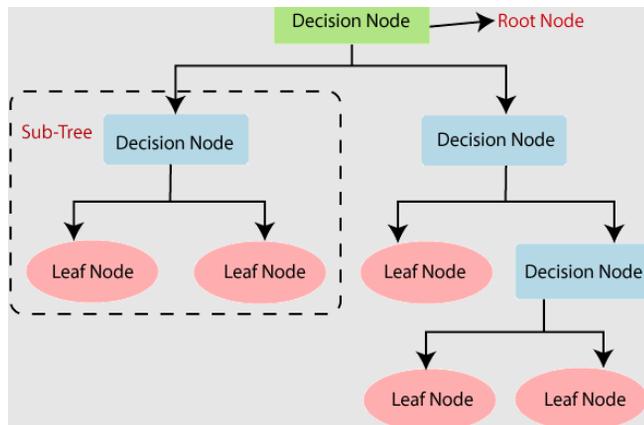
전통적인 기계학습은 크게 지도학습(Supervised learning), 비지도학습(Unsupervised learning)의 두 가지로 나눌 수 있다. 지도학습은 컴퓨터의 학습을 위해 정답이 있는 데이터를 제공하는 방식이다. 즉, 지도학습을 위한 데이터는 입력값에 대한 결과값(레이블)이 주어지며, 분류 및 회귀분석이 지도학습을 사용하는 대표적 유형이다. 대표적으로 일반화 선형 모델(Generalized linear models), 의사결정나무(Decision tree), 랜덤 포레스트(Random forest)가 있다. 반면 비지도학습은 정답이 없는 데이터를 기반으로 컴퓨터를 학습시킨다. 일

반적으로 비지도학습은 학습용 데이터를 비슷한 특징에 따라 군집화하는 등의 방법을 통해 학습을 수행한다. 대표적으로 k-평균 군집분석(k-means clustering), 주성분 분석(Principal component analysis) 등이 있다.

강화학습(Reinforcement learning)은 기계학습의 한 부분이나, 기존의 지도학습과 비지도학습과는 다른 분류방식이다. 지도학습과 비지도학습이 학습용 데이터의 차이에 따른 분류였다면, 강화학습은 학습 방식에 차이가 있다. 강화학습은 데이터를 통한 경험을 통해 성능을 향상시킨다. 강화학습 알고리즘은 현재 상태에서 취하는 행동에 따른 보상이 주어지며, 이같은 보상을 최대화할 수 있도록 학습을 수행한다.

1) 의사결정나무

의사결정나무는 지도학습 모델 중 하나로 분류와 회귀분석에 모두 사용가능하다. 의사결정나무를 직관적으로 이해한다면, 여러 차례의 질문을 통해 정답을 좁혀나가는 ‘스무고개’를 생각할 수 있다. 이처럼 의사결정나무는 여러 가지 규칙을 순차적으로 적용하는 방식으로 데이터를 처리한다. 또한, 순차적으로 적용되는 규칙들을 배열한 모양이 나무와 같다고 하여 의사결정나무라 지칭한다.



[그림 2-13] 의사결정나무의 일반적 형태

출처 : DEVOPS 홈페이지, <https://www.devops.ae/decision-tree-classification-algorithm/> (검색일 : 2022.06.15.)

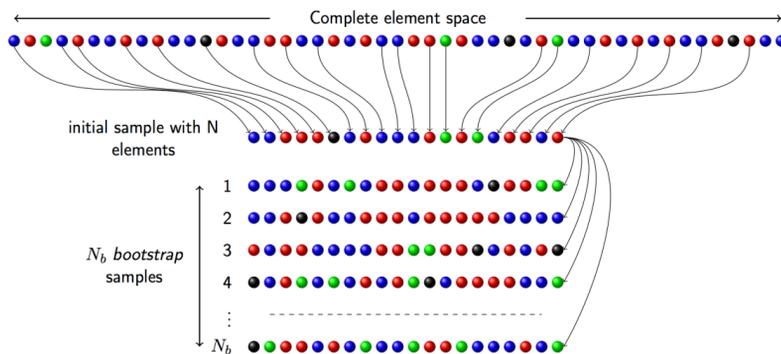
의사결정나무는 거꾸로 뒤집은 나무의 모양과 유사하다. 가장 위에 있는 뿌리 노드로부터 가지가 뻗어나와 다른 노드와 연결되며, 마지막 노드들이 잎처럼 달린 형상이다. 의사결정나무에서 각 노드는 질문이나 질문의 결과로 나온 정답을 말한다. 뿌리 노드와 의사결정 노드는 질문, 입력자료 및 정답이 포함되며, 가장 마

지막단의 잎 노드는 정답만 포함한다. 한 번의 질문 노드는 일반적으로 변수 영역을 2개로 구분할 수 있도록 2개의 노드로 분리된다. 즉, 의사결정나무의 입력데이터는 노드를 거칠 때마다 의사결정규칙에 따라 출력값이 결정되고, 이 출력값은 다시 다음 노드들의 의사결정규칙에 따라 최종 출력값이 결정된다.

의사결정나무 모델은 이같이 상대적으로 단순한 방식으로 정답을 추론한다는 장점이 있으나, 다른 지도학습 모델들에 비해 상대적으로 오버피팅 가능성이 높은 등 성능이 떨어진다는 단점이 있다. 이같은 단점을 극복하기 위해 의사결정나무를 기저모델로 사용하는 앙상블 방식의 랜덤포레스트가 개발되었다.

2) 랜덤포레스트

랜덤포레스트는 이름 그대로 의사결정나무들이 모여 숲을 이룬 모형이다. 랜덤포레스트는 기존 의사결정나무의 단점을 극복하기 위해 다수의 의사결정나무 모형을 만들어 복잡적으로 이용함으로써 성능을 향상시키고자 하였다. 그러나 의사결정나무 모형은 하나의 데이터셋에서 하나밖에 만들 수 없다. 하나의 학습용 데이터셋으로부터 여러 개의 의사결정나무 모형을 만들기 위해 기존 학습용 데이터셋에서 랜덤하게 데이터를 추출하는 부트스트랩 샘플링(Bootstrap sampling) 과정을 수행한다. 즉, 하나의 학습용 데이터셋에서 중복을 허용하여 학습용 데이터셋보다 작은 크기의 샘플을 여러개 추출한다.

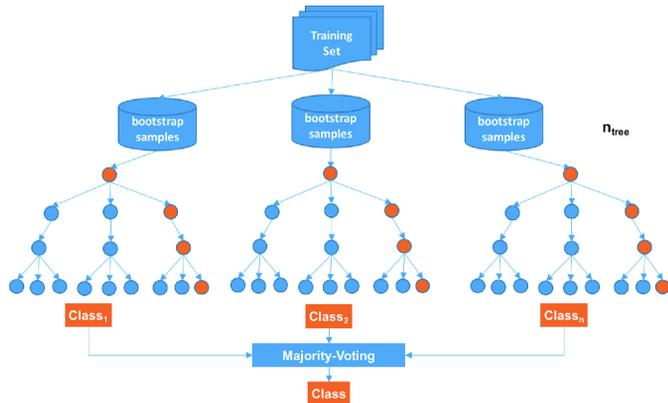


[그림 2-14] 부트스트랩 샘플링 개념

출처 : Let the Machine Learn 홈페이지, <https://yashuseth.wordpress.com/2017/12/02/bootstrapping-a-resampling-method-in-statistics/> (검색일 : 2022.06.15.)

추출된 각 샘플을 통해 각 샘플에 맞는 의사결정모형을 만들어낼 수 있다. 추출하는 샘플의 개수는 사용자에게 의해 조절 가능하며, 샘플의 개수만큼 의사결정나무가 만들어진다. 만들어진 랜덤포레스트 모형의 구동시 각각의 의사결정 나무들

은 데이터에 대한 분류 및 회귀 결과를 도출한다. 랜덤포레스트는 이같은 여러 개의 결과물들을 취합하여 투표를 실시하여 가장 많이 득표한 결과물을 최종 결과물로 선택한다. 즉, 100개의 의사결정나무가 1~5까지 5개의 결과를 예측했으며, 각각 10, 20, 20, 20, 30개의 의사결정나무의 예측 결과라면, 5를 최종 결과물로 선택한다.



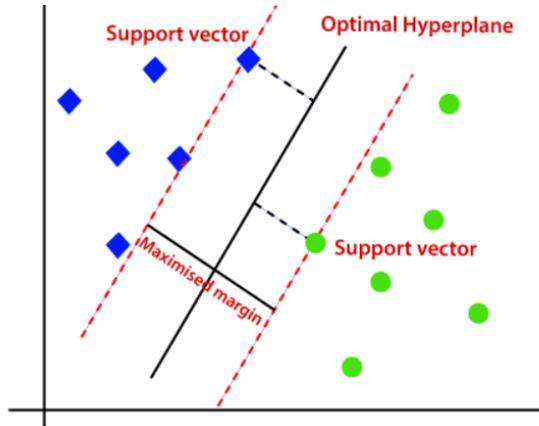
[그림 2-15] 랜덤포레스트 기반의 앙상블 분류모델 구축 개념

출처 : Medium 홈페이지, <https://medium.com/@uma.bollikonda/random-forest-machine-learning-algorithm-401bdcd7a0b8> (검색일 : 2022.06.15.)

랜덤포레스트는 의사결정나무의 단점인 오버피팅을 극복하기 위해 여러 개의 의사결정나무를 사용한다. 생성된 의사결정나무들 중 일부 의사결정나무는 오버피팅될 수 있으나, 생성되는 의사결정나무 수 자체를 늘림으로써 오버피팅의 영향을 최소화시킨다.

3) 서포트 벡터 머신

서포트 벡터 머신은 분류를 위한 기준 선인 결정 경계(Decision boundary)를 정의하는 모델로, 다른 지도학습 기반의 모형과 비교하여 데이터 분류 결과가 정확하고, 특히 이진데이터 분류에 강점을 나타낸다. 서포트 벡터 머신은 최적의 결정 경계(Optimal hyperplane)를 도출하기 위해 데이터 학습을 수행한다. 결정 경계와 가장 가까운 파란색 및 초록색 점까지의 거리를 마진이라고 하며, 최적의 결정 경계는 이 마진을 최대화하는 방향으로 생성된다. 또한, 결정 경계와 평행하게 그려지며, 가장 가까운 점을 지나는 직선을 서포트 벡터라고 한다. 선형의 결정 경계를 이용하여 분류가 불가능한 경우 polynomial, sigmoid 등 곡선을 이용하기도 한다.

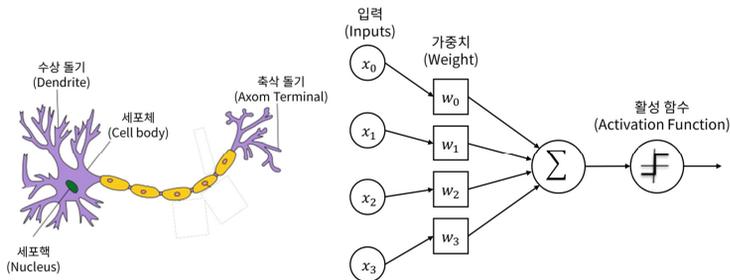


[그림 2-16] 서포트 벡터 머신의 기본 개념

출처 : Medium 홈페이지, <https://medium.com/@viveksalunkhe80/support-vector-machine-svm-88f360ff5f38> (검색일 : 2022.06.15.)

4) 인공신경망

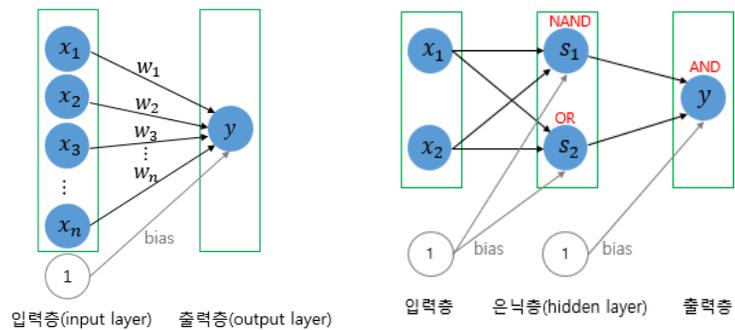
인공신경망은 인간의 뉴런 구조에 착안하여 만든 기계학습 모형이다. 가장 기초적인 인공신경망인 퍼셉트론은 1957년 Rosenblatt에 의해 고안된 알고리즘으로, 다수의 입력신호를 받아 하나의 신호를 출력하는 뉴런의 수상돌기 및 축삭돌기의 구조를 모방하고 있다. 퍼셉트론에서는 입력데이터가 뉴런의 입력신호의 역할을 하며 가중치가 뉴런의 신호전달 역할을 수행한다. 즉, 입력된 데이터는 퍼셉트론의 가중치 연산을 수행하여 최종 결과물을 출력한다. 하나의 입력 및 출력층으로 구성된 퍼셉트론을 단층 퍼셉트론이라고 하며, 단층 퍼셉트론은 선형 분리가 가능한 문제에만 적용가능하다는 한계가 있다. 이같은 단층 퍼셉트론의 한계를 극복하기 위해 다층 퍼셉트론이 개발되었다.



[그림 2-17] 뉴런을 모방한 퍼셉트론의 개념

출처 : Compmath 홈페이지, <https://compmath.korea.ac.kr/appmath2021/Perceptron.html> (검색일 : 2022.06.15.)

다층 퍼셉트론은 기존 단층 퍼셉트론이 갖는 입력층과 출력층 사이에 은닉층 (Hidden layer)을 추가하여 구성한다. 은닉층을 추가함으로써 퍼셉트론은 더 복잡한 문제에 대응 가능해진다. 은닉층이 추가된 다층 퍼셉트론은 기존의 단층퍼셉트론이 사용하는 방법으로는 모형을 학습시킬 수 없다. 단층 퍼셉트론은 학습 데이터를 입력층부터 출력층까지 순전파(Feedforward) 시키고, 출력데이터와 실제 결과값을 비교함으로써 가중치를 변경하는 등 모형을 학습시킬 수 있다. 반면, 다층 퍼셉트론은 은닉층에서 계산되는 출력값의 정확도를 비교할 기준이 없어 최종 출력층의 오차값을 은닉층으로 되돌리는 역전파(Backpropagation) 과정을 수행한다. 다층 퍼셉트론은 이와 같은 순전파와 역전파를 이용하여 모형을 학습시키며, 더 복잡한 문제를 해결하기 위해 은닉층의 개수를 2개 이상으로 증가시키는 경우 이를 심층신경망 혹은 딥러닝이라고 한다.



[그림 2-18] 단층 퍼셉트론과 다층 퍼셉트론

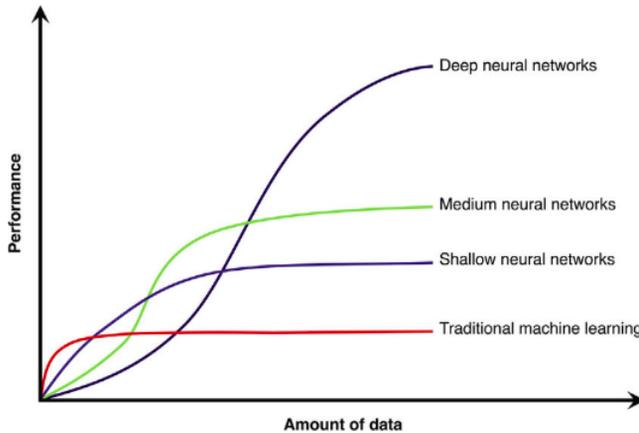
출처 : WikiDocs 홈페이지, <https://wikidocs.net/24958> (검색일 : 2022.06.15.)

5) 딥러닝

딥러닝은 인공신경망에 여러 개의 은닉층을 추가하여 더 깊은 구조를 구축한 심층 신경망(Deep neural network) 기반의 기계학습 방법이다. 즉, 다층 퍼셉트론이 확장이라 볼 수 있다. 이같은 딥러닝 기술을 여러 개의 신경망을 연결시켜 기존에 해결하지 못했던 더 복잡한 문제를 해결하는데 사용되기도 한다. 신경망의 깊이가 깊어질수록 컴퓨터의 연산능력 요구량이 높아져 사용에 어려움이 있었으나, 4차산업혁명 기술의 발달과 함께 하드웨어 성능이 증가하여 딥러닝 기반의 기술개발이 활성화되고 있으며, 다른 기계학습 모형과 비교할 때 더 높은 정확도와 성능을 나타내고 있다.

다만 딥러닝은 동일한 수준의 모형 학습을 위해서도 더 많은 데이터와 연산을 필요로 한다는 단점이 있다. 과거 딥러닝이 최근 들어 적용되고 있는 이유도 이와

같다. 기존의 머신러닝 기법들은 모형이 일정 수준의 성능을 갖추기 위해 필요한 데이터량이 많지 않았으며, 일정량 이상의 데이터를 넘어설지라도 모형의 성능이 크게 향상되지 않았다. 그러나 딥러닝은 기존 머신러닝 기법과 비교할 때 훨씬 많은 양의 데이터를 요구하고 있다. 딥러닝에 필요한 방대한 규모의 데이터 수집을 위해 데이터 큐레이션의 중요성이 높아지고 있으며, 데이터 큐레이션을 위한 산업이 추가적으로 발생하고 있는 상황이다.



[그림 2-19] 학습을 위해 필요한 데이터량과 모형의 성능사이 관계

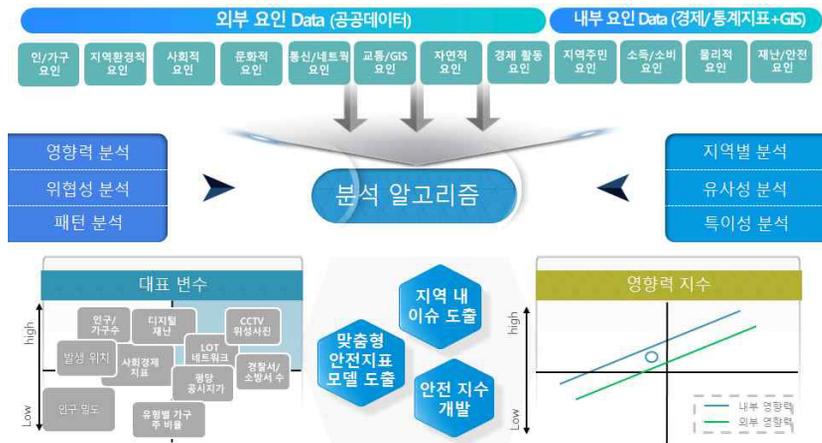
출처 : Medium 홈페이지, <https://medium.com/@tanmayshimpi/difference-between-ml-and-deep-learning-with-respect-to-splitting-of-the-dataset-into-375d433ee2c8> (검색일 : 2022.06.15.)

딥러닝이 갖는 또 하나의 단점은 히든레이어의 존재에서 발생한다. 인공신경망을 심층신경망으로 만들어주는 히든레이어는 딥러닝 분석 과정의 의미와 결과에 대한 원인을 이해하기 어렵게 만든다. 입력자료와 출력자료 사이에 수천개 이상의 매개변수가 존재하며, 매개변수와 입력값 간의 관계가 일대일 관계가 아니기 때문에 원인과 결과에 대한 명확한 해석이 불가능하다. 즉, 입력자료에 대한 출력결과가 나왔을 때 연구자의 직관에 따라 인과관계를 유추할 수는 있으나, 실제 모형의 학습결과를 이용하여 원인이 무엇인지 설명하기는 어렵다.

4. 빅데이터 기반 재난·재해 예측 연구

1) 빅데이터 기반 재난재해 연구 동향

김태형(2015)은 빅데이터를 활용하여 서비스 수요자의 니즈가 반영된 재난안전 분야 스마트시티 서비스를 제공하고자 하였다. 이를 위해 데이터 매쉬업을 통해 공공데이터와 민간데이터를 결합한 예측 모델을 수립하고 가구별 재난안전 유형을 파악하여 수요자 맞춤형 서비스를 제공하고자 하였다.



[그림 2-20] 데이터 매쉬업을 통한 재난안전 유형 판별 알고리즘

출처 : 김태형(2015). 수요자 중심의 Smart City 구축을 위한 빅데이터 활용 방안: 재난안전 분야를 중심으로. 한국지역정보학회지 18(2), p.58.

인구 및 지리정보 데이터와 같이 통계청 등에서 발표하는 공공데이터와 함께 기업·소상공인 DB, 신용카드 매출 데이터 등 민간데이터를 활용하여 지역 상권의 특성, 소득소비 수준 등 가구를 유형화 할 수 있는 변수를 설정하고 각 유형별 특성을 파악했다.

[표 2-9] 데이터 매쉬업을 위한 활용 데이터

구분	데이터	원천 데이터 보유 기관
인구통계학 데이터	주민등록인구현황	안전행정부, 행정자치부, 통계청
	주택 유형 및 가격	통계청, 한국감정원
	경제활동 인구통계	통계청, 신용평가사
권역 및 지리 정보	전국 블록 및 유형 정보	통계청, 국토교통부
소득 데이터	소득분위	통계청
거래/소비패턴	거래 패턴	신용평가사, 카드사, 통신사
	고객 정보 (성, 연령, 주소의 비식별 정보)	

출처 : 김태형(2015). 수요자 중심의 Smart City 구축을 위한 빅데이터 활용 방안: 재난안전 분야를 중심으로. 한국지역정보학회지 18(2), p.59.

데이터 매쉬업을 통해 가구 유형별로 재난안전 유형이 추가되어 보다 정확한 가구별 재난안전 유형 정보를 확인할 수 있다. 이를 활용하여 재난안전맵을 구축할 경우 재난 발생을 예측하고 예방을 위한 수요자 중심의 스마트도시 서비스를 제공할 수 있을 것으로 예상된다.



[그림 2-21] 재난안전 맵 활용(예)

출처 : 김태형(2015). 수요자 중심의 Smart City 구축을 위한 빅데이터 활용 방안: 재난안전 분야를 중심으로. 한국지역정보학회지 18(2), p.63.



[그림 2-22] 수요자 중심의 안전도시 에코시스템(예)

출처 : 김태형(2015). 수요자 중심의 Smart City 구축을 위한 빅데이터 활용 방안: 재난안전 분야를 중심으로. 한국지역정보학회지 18(2), p.65.

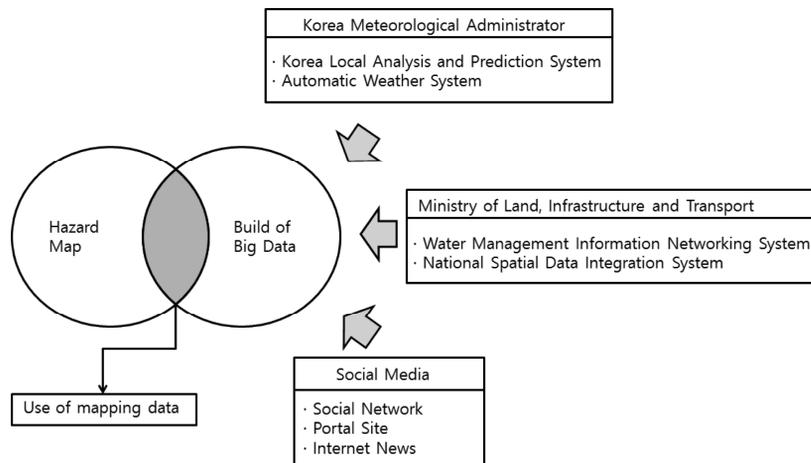
김영민 외(2015)는 기상에 의한 자연재해 분석을 위해 관련 분야에서 활용되고 있는 데이터 마이닝과 기계 학습 기술에 대한 고찰을 진행하였다. 기상데이터에 대한 예측, 자연재해와 연관되는 지형 분석, 지진이나 산불과 같은 자연재해 분석 등에 빅데이터 및 기계 학습이 활용되고 있다. 기상 및 자연재해 예측에는 SVM, ANN 등이 주로 사용되며, 퍼지 시스템, 유전자 알고리즘, 베이지안 방법론 또한 사용된다. 지형 분석에는 이미지 분류, 클러스터링이 활용되었다.

임현택 외(2016)는 빅데이터를 활용하여 침수 위험에 대응한 재해지도를 보다 효율적으로 작성하는 방법에 대해 연구하였다. 기존 재해지도의 수정·보완 주기는 5년 또는 10년 단위이며, 이미지 또는 문서 파일 등으로 저장되는 등 활용도가 낮은 단점이 있었다. 이에 저자는 공공에서 제공하는 공간데이터와 행정데이터 그리고 민간에서 제공하는 데이터를 함께 활용하여 재해지도 구축에 사용하여 이러한 단점을 극복하고자 하였다. 실시간 민간 정보들이 공간 및 속성정보와 중첩되어 작성된 재해지도를 통해 침수원인 파악, 피해 예측, 대피 경로 탐색 등에 활용될 수 있을 것으로 보았다.

[표 2-10] 침수 대응 재해지도 구축 활용 데이터

구분		데이터 종류
공공	공간데이터	(국가공간데이터통합체계) 수치지도, 도로경계, 지적도 등 주제도 (기상청) 강우이력자료, 실시간 강우자료
	행정데이터	(국토교통부) 한강홍수통제소의 시간대별 수위정보, 침수 발생지역의 침수 순간의 수위정보, 지역단위의 예·경보 현황 자료
민간		실시간 SNS(트위터 등), 웹뉴스, 블로그, 포털사이트 등 소셜미디어 정보

출처 : 임현택 외(2016) 참고하여 연구진 작성.



[그림 2-23] 빅데이터 스키마의 구축 및 적용(Building and Applying Scheme of Big Data)

출처 : 임현택 외(2016). 재해지도 활용성 증대를 위한 빅데이터 구축 및 적용 방안. 한국방재학회논문집 16(2), p.211.

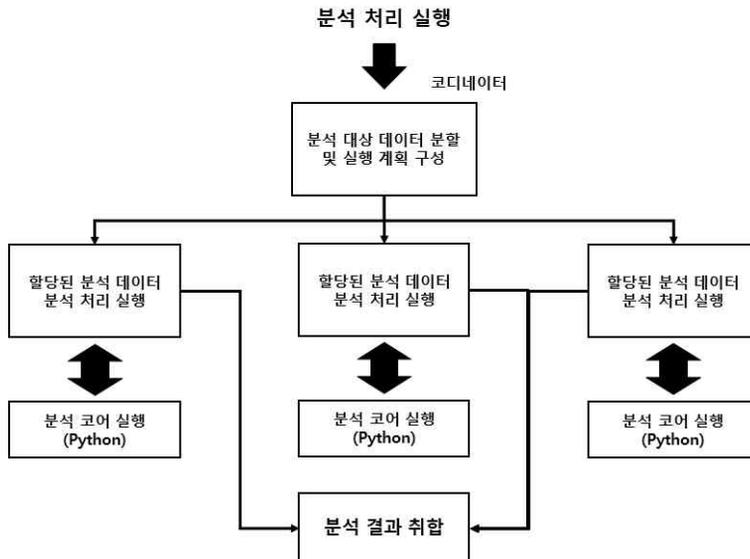
이동규(2016)는 미국과 영국 등 선진국과 국내의 빅데이터를 활용한 재난관리 시스템을 조사·분석 하였다. 빅데이터는 재난의 발생 전 예방에서부터 발생 시의 대응, 그리고 발생 후의 복구까지 모든 과정에서 활용될 수 있어 효과적인 재난관리가 가능하다. 다만 이를 위해서는 ‘통합상황실’ 등을 활용한 정보의 통합 운영이 필수적이다.

[표 2-11] 빅데이터 활용 재난관리 시스템

국가	기관	시스템	수준	
			데이터 분석	재난관리 단계
한국	국립 재난안전연구원	스마트 재난상황실	예측	대응
		도시내수 침수 전조감지 시스템	예측	대비
미국	국립해양대기청	기상경보체계	예측	대비
	지질조사국	지질분석 시스템	예측	예방
		트위터 지진감지기	서술	대비
영국	국무조정실	국가 및 지역 재난안전 리스크 관리	진단	예방
	환경청	실시간 홍수경고지도	예측	대비

출처 : 이동규(2016). 빅데이터 기반의 재난관리 시스템 운용 방향에 대한 예비적 고찰 - 미국, 영국, 한국의 사례 비교를 중심으로 -. 한국위기관리논집 12(1), p.28.

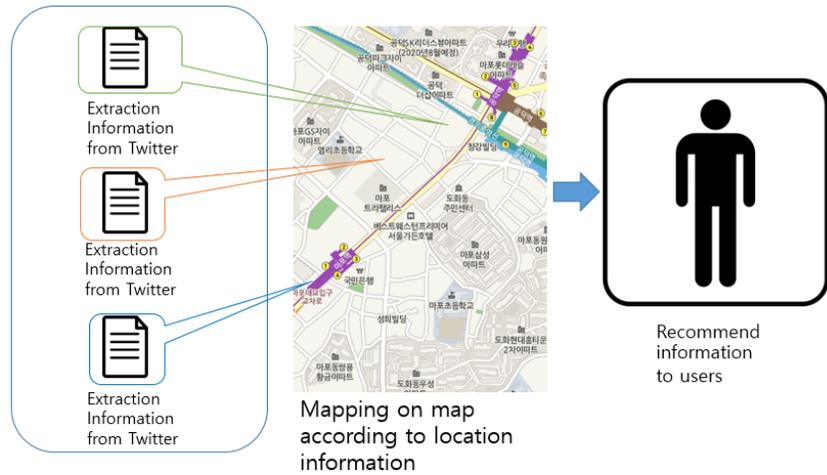
채수성 외(2018)는 데이터를 수집하고 머신러닝 기술 등을 활용해 분석하여 재난 상황의 전 과정에서 의사결정을 지원하는 시스템을 구축 제안한다. 이는 복합 재난 등이 발생했을 경우, 다양한 유형의 데이터를 활용한 기계학습 등을 빠르게 진행하여 기존의 시스템에서는 어려웠던 분석 대상의 규모가 거대한 빅데이터 분석을 가능하게 한다. 기존의 재난 위험도 분석 시스템인 ERGO를 기반으로 분산병렬 처리를 통해 대규모의 데이터를 기계학습으로 정확히 분석하는 시스템으로 재난 및 재해 상황시 신속정확한 의사결정을 지원할 수 있을 것으로 제안하고 있다.



[그림 2-24] 분산병렬 재난 위험도 분석 처리 흐름도

출처 : 채수성 외(2018). 빅데이터 기반 재난 재해 위험도 분석 프레임워크 설계 및 구현. 디지털 콘텐츠학회논문지 19(4), p.775.

신봉희와 전해경(2018)은 스마트폰의 보급과 SNS의 이용 증가 추세를 근거로 재난 발생시 SNS를 통해 실시간으로 비정형의 빅데이터 정보를 수집할 수 있을 것으로 보았다. 이에 트위터(twitter)라는 SNS를 이용하여 재해 정보를 수집하는 시스템을 구축하고 유효성을 평가하였다.

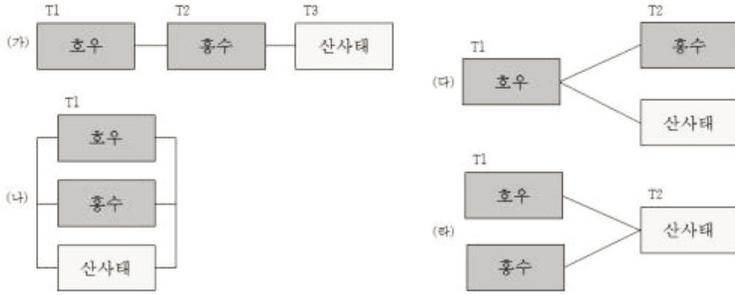


[그림 2-25] 필요 정보 추출 시스템의 흐름

출처 : 신봉희 외(2018). 빅 데이터를 이용한 재해 정보 지원에 관한 연구. 한국융합학회논문지 9(8), p.27.

해당 시스템은 해시태그를 이용해 재난재해 정보를 수집하고 정보의 발생위치를 함께 수집하여 지도상에 맵핑한다. 시범 구축한 시스템은 정보의 양이 많을 때 속도가 느려지는 것을 제외하고 수집·맵핑된 재난 정보의 정확도는 높아 SNS를 통해 생산되는 빅데이터를 활용하여 재난재해 정보의 수집 및 효과적인 분석의 가능성을 제시한다.

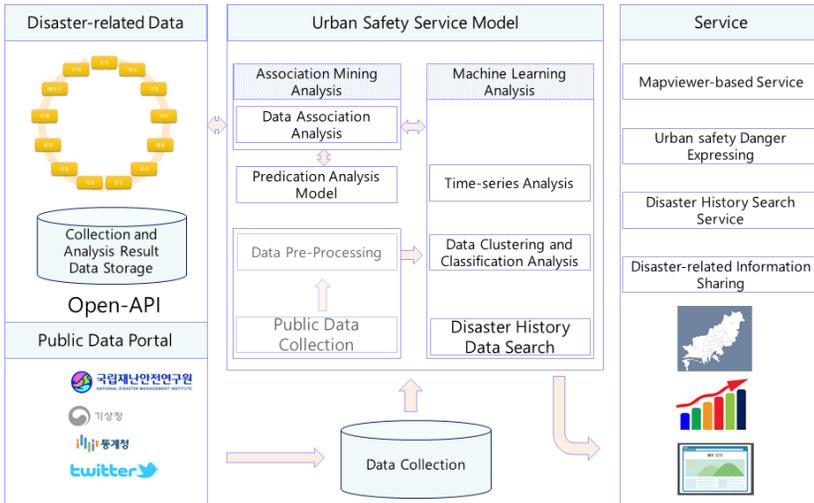
나유경 외(2019)는 다양한 피해 양상이 나타나는 복합재난에 대응하기 위해 온라인 뉴스 데이터를 활용하여 복합재난 발생 지역 및 유형 분석을 진행하였다. 크롤링을 통해 수집한 뉴스 데이터를 행정구역 지명, 재난 용어 빈도수 측정 등의 전처리 과정을 거쳐 재난 발생일, 발생 지역, 재난 종류 등을 추출하여 재난별, 지역별 발생현황을 분석가능하다. 국내 기후환경을 고려하여 강우(호우, 홍수 또는 호우와 홍수)에 의한 산사태 발생 시나리오를 사례로 선정하여 분석을 진행한 결과, 가정한 복합재난 시나리오의 발생가능성을 확인 가능했고, 발생 빈도가 높은 지역 및 시기의 파악이 가능했다. 해당 분석을 통해 향후 국내의 실정에 맞는 복합재난 유형 분류에 도움을 줄 수 있으며, 지역별 및 시기별 발생재난에 대한 통찰력을 주어 복합재난의 예방 및 대응에 기여할 수 있을 것으로 기대했다.



[그림 2-26] 호우 및 홍수로 인한 산사태 발생 시나리오

출처 : 나유경 외(2019). 빅데이터를 이용한 복합재난의 발생 지역 및 유형 분석. 국토지리학회지 53(4), p.380.

정명균 외(2020)는 기존의 지역안전지수에 재난 이력정보, 실시간 기상 데이터, SNS 데이터 등 빅데이터를 기계학습으로 분석하여 반영한 예측모델을 생성하고, 예측모델의 안전지수 결과를 지도에 맵핑하여 제공하는 시스템을 제안했다.



[그림 2-27] 안전지수 서비스 모델 구조

출처 : 정명균 외(2020). 빅데이터 분석을 활용한 재해 분야별 안전지수 서비스 모델 연구. 한국재난정보학회논문집 16(4), p.686.

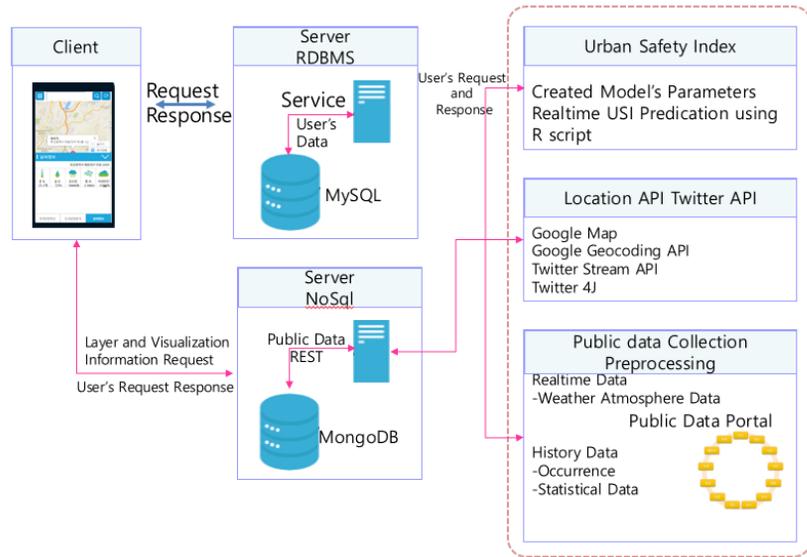
서비스 이용자의 위치를 기반으로 해당 지역의 안전사고 발생과 관련한 정보와 SNS 데이터를 수집해 기계학습 알고리즘을 활용하여 연관성 및 관계성을 분석한다. 기계학습 알고리즘 중 다항로지스틱회귀분석, SVM (Support Vector Machine), 신경망기법을 적용하여 화재, 교통, 범죄, 자연재해에 대한 지역 안전지수와 기상대기 데이터의 상관요소를 확인하고 연관성 모델 분석을 실시했다. 예측 모델의 결과를 기반으로 이용자가 기상 정보 등 실시간 데이터가 반영된 도시안전지수를 제공받을 수 있는 시스템을 제안하였다.

[표 2-12] 재해 분야별 안전지수 모델 분석 결과

분야	속성	예측 확률*		
		M	S	N
화재	기온, 풍속, 미세먼지, 습도, 풍향	36%	44%	35%
교통	기온, 최고기온, 강수량, 풍속, 풍향, 미세먼지, 습도	32%	45%	40%
범죄	기온, 최고기온, 습도	26%	32%	30%
자연재해	최저기온, 풍속, 강수량, 습도	30%	33%	29%

* 주: M (다항로지스틱회귀분석), S (SVM), N (신경망기법)

출처 : 정명균 외(2020). 빅데이터 분석을 활용한 재해 분야별 안전지수 서비스 모델 연구. 한국재난정보학회논문집 16(4), p.687.



[그림 2-28] 이용자 위치기반 중심 실시간 데이터 반영 안전지수 표출 시스템 설계도

출처 : 정명균 외(2020). 빅데이터 분석을 활용한 재해 분야별 안전지수 서비스 모델 연구. 한국재난정보학회논문집 16(4), p.688.

2) 빅데이터 기반 화재 연구 동향

박은석과 민세홍(2019)은 기존 화재관련 정보시스템이 데이터의 부정확성과 빅데이터 분석을 위한 처리의 복잡성 등의 문제를 가지고 있으며, 소방청 국가화재정보시스템의 화재네비게이터의 갱신 주기가 연 1회 임을 지적했다. 이에 실시간 화재발생 예측 플랫폼을 구축하기 위해 화재통계 데이터를 분석하여 국내 화재 발생 패턴을 확인하고 화재영향인자를 도출했다.

국내 화재발생 패턴은 부주의와 전기적 요인에 의한 경우가 많았으며, 이에 대한 화재영향인자를 데이터의 연관성, 지속성, 호환성, 생산성을 기준으로 평가하여 도출했다. 우수(3)/양호(2)/보통(1)/미흡(0)의 4단계 평가 점수의 합이 9점 이상

인 경우 화재영향인자로 선정했으며, 전기 사용량, 가스 사용량, E-지방지표(인구, 스트레스 인지율, 자살율 등), 건축물 노후도, 노령화 인구, 토지이용률, 가구 및 소득분포의 7개 데이터가 선정되었다.

각 영향인자의 데이터를 적용한 화재예측플랫폼은 화재발생확률을 예측하여 실시간 예방에 도움을 줄 것으로 예상하고 있다.

[표 2-13] 화재영향인자 데이터 평가 결과

화재영향인자 데이터	연관성	지속성	호환성	생산성	합계
전기 사용량	3	3	3	2	11
가스 사용량	3	3	3	2	11
E-지방지표	2	2	3	2	9
건축물 노후도	3	3	3	3	12
노령화 인구	2	3	3	3	11
행정구역별 토지이용률	3	2	3	3	11
가구 및 소득분포	3	2	3	2	10
금연구역 현황	3	0	3	0	6

출처 : 박은석 외(2020). 빅데이터 활용을 위한 화재인자 표준화에 관한 연구. 한국방재학회논문집 19(4), p.148.



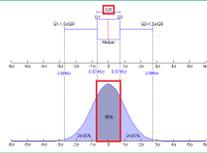
[그림 2-29] 화재예측플랫폼 예시

출처 : 박은석 외(2020). 빅데이터 활용을 위한 화재인자 표준화에 관한 연구. 한국방재학회논문집 19(4), p.148.

김도형과 조병완(2021)은 재난여건 다변화에 따른 현장의 초기 대응의 어려움을 극복하기 위해 화재관련 통계 데이터의 변수간 선형 상관성 분석, 머신러닝 알고리즘 활용 분석을 실시하고, 화재현장 위험도의 등급화, 초기 정보를 이용한 화재현장 위험도 예측 모델을 개발했다.

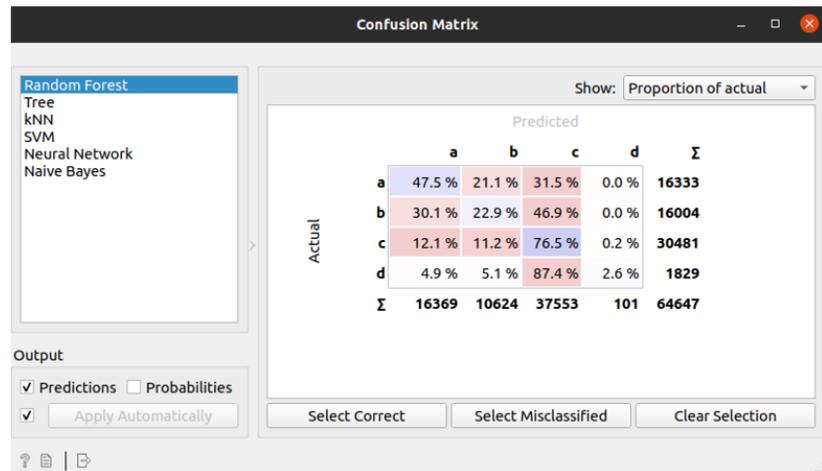
화재에 의한 피해를 나타내는 인명피해, 재산피해, 이재민수, 이재세대수의 통계 데이터를 이용하여 0~4의 범위로 화재위험도를 등급화 했다.

[표 2-14] 화재현장 위험도의 정의 및 등급화

화재 위험지수 산정식	4분위수 개념	화재위험도 등급	데이터 수
$0.4 \times (\text{사망자} \times 0.8 + \text{부상자} \times 0.2) + 0.4 \times \text{재산피해} + 0.2 \times \text{이재피해}$		A (위험-하)	34,475
		B (위험-중)	18,804
		C (위험-상)	11,368

출처 : 김도형 외(2021). 빅데이터 분석을 활용한 초기 정보 기반 화재현장 위험도 예측 모델 개발 연구. 한국재난정보학회논문집 17(2), p.251.

예측모델에 적용할 머신러닝 알고리즘은 Random Forest를 사용했으며, A 등급 47%, B 등급 22.9%, C 등급 76.5%의 정확도를 보였으며, 정확도의 향상을 위해서는 인명피해와 재산피해 데이터의 편향에 의한 것으로 보여 향후 데이터 정합과 결손에 대한 개선이 필요한 것으로 나타났다.



[그림 2-30] 머신러닝 알고리즘 분석 결과

출처 : 김도형 외(2021). 빅데이터 분석을 활용한 초기 정보 기반 화재현장 위험도 예측 모델 개발 연구. 한국재난정보학회논문집 17(2), p.252.

5. 소결

1) 국내외 재난·재해 빅데이터 기반 예측·예방·대응 시스템

국가는 재난 및 재해 예방·관리·복구를 위해 다양한 위기관리 시스템을 구축하여 운영하고 있으며, 관련 시스템은 다양한 재난·재해 데이터를 기반으로 운영되고 있다. 국내 재난·재해 예방 및 대응 주무부처는 행정안전부 재난안전관리본부가 총괄하고 있으며, 재난·재해 예방 및 대응을 위해 대표적으로 국가재난관리정보시스템 및 안전정보통합공개시스템을 운영하고 있다.

「재난 및 안전관리 기본법」에 국가재난관리정보시스템은 재난관리업무, 국민재난안전, 모바일재난관리, 안전디딤돌의 4가지 시스템을 운영 중이며, 국가 재난 관리에 관한 주요 인력 및 물자의 체계적인 관리와 재난·재해 발생 시 상황의 신속한 전파 및 체계적인 복구 수행을 목적으로 하고 있다.

안전정보통합공개시스템은 집중적으로 관리되어야 하는 시설 및 건축물에 대한 안전점검 및 진단 등의 정보를 통합적으로 공개하고 있으며, 2023년 최종구축을 목표로 하고 있다. 주요 공개정보는 각 시설 및 건축물에 대한 점검종료, 점검일, 점검 보고서이며, 시설 및 건축물 유형에 따라 개별부처에서 안전관리를 담당하고 있기 때문에 공개정보는 시설 및 건축물 특성에 따라 달라지고 있다. 특히 대국민 안전을 위해 6개 분야(재난·치안·교통·보건·생활·시설)에 대한 생활안전 지도를 제공함으로써 국민들이 쉽게 안전관련 정보를 인지할 수 있도록 한다.

사회재난인 화재와 범죄는 개별적인 시스템을 구축하여 운영하고 있다. 화재는 국가화재정보시스템을 통해 지역별 화재발생위험도에 관한 정보를 제공하여 사전에 화재발생을 예방하고, 화재예방정책 수립 및 대응자료로 활용되고 있다. 이외 화재유형 및 건축물유형별 화재발생 등 다양한 화재사고 통계를 제공하고 있다. 경찰청에서 운영하는 지리적 프로파일링 시스템은 범죄예방을 위한 범죄위험도 측정을 주로 수행하고 있으며, 범죄발생 건수, 주거인구 및 유동인구 수, 유흥업소, 가로등·CCTV 등의 자료를 바탕으로 범죄발생 위험이 높은 지역을 지방경찰청 및 관할 경찰서에서 파악하고 범죄를 사전에 예방하고자 노력하고 있다.

국외의 화재발생 예측 시스템은 우선적으로 대규모 화재 발생 모니터링을 통한 화재 발생 및 확산 예측 시스템을 주를 이루고 있다. FIRECAST는 위성관측을 통해 대규모 화재가 발생할 것으로 예측되는 생태계 위험요인을 모니터링하여 고객에게 제공하고 있으며, 고객들은 이를 통해 전략적 토지 이용 및 정책을 수행하고 있다. 미국 내무부(Department of the Interior)는 Fire Danger Forecast를

운영하고 있으며, 미국 내 잠재적 화재 위험성을 모니터링하여 디지털 지도를 자동적으로 생성하는 프로세스를 유지하고 있다. 특히 화재발생 관련 각종 지수·지표를 제공하고 있으며, 화재위험지수, 화재발생가능성지표, 화재확산가능성지표, 화재위험예측 식생지수, 지역별 화재예측그래프를 제공함으로써 화재예방, 대응 및 확산 방지를 위한 정책적 판단을 지원하고 있다.

해외 건축물 화재위험도 예측모델로는 미국 카네기멜론 대학에서 구축한 Metro21 프로젝트가 있다. 특정건축물을 대상으로 6개월 내 화재 발생 가능성을 예측하는 모델로 화재발생데이터, 건축물 가치, 건축물 유형, 토지정보, 소음 및 시설 위반사항 등의 데이터를 활용하여 화재발생을 예측하고, 화재위험도를 1~10점의 리커트 척도로 보여주고 있다.

화재 이외 재난·재해 예측시스템은 인도 및 방글라데시 일부 지역을 대상으로 구글이 운영하는 홍수예측시스템이 있다.

국내외 재난·재해 관련 빅데이터 예측 시스템 고찰을 통해 도출한 시사점은 다음과 같다. 우선 재난·재해 예측을 통한 예방을 위해서는 관련 빅데이터의 수집·가공 및 분석을 위한 시스템 구축이 필요하다. 모든 재난·재해의 발생을 예측하는 것은 어렵지만 관련 데이터의 수집 및 분석을 통해 일정부분 예측가능하며, 이를 통해 재난·재해에 대응하기 위해 우선적으로 예방정책을 도입해야 하는 지역 및 건축물을 도출할 수 있다. 또한 재난·재해 발생에 대한 신속한 대응 및 조치와 피해최소화를 위해 정책수행자 및 일반국민이 이해하기 쉽게 관련 자료를 제공해야 한다. 재난·재해 예측시스템은 다양한 발생인자를 빅데이터 분석 방법론 및 복잡한 통계방식을 통해 위험도가 높은 지역 및 건축물을 도출하는 것이기 때문에, 프로세스 전반을 이해하는 것은 어려운 부분이 있다. 이에 이를 지수·지표화하거나 안전지도 등을 통해 국민이 알기 쉽게 제공하여 미리 대비할 수 있도록 하는 것이 중요하다. 마지막으로 건축물의 재난·재해 예측모델 구축을 위해서는 건축물 속성정보가 중요하다. 건축물 유형, 연면적, 높이, 노후도, 점점이력 등이 주요 예측요인으로 활용되기 때문에, 건축행정 데이터 정제 및 정확도를 높이는 사업이 수행될 필요가 있다.

2) 빅데이터 분석 방법론 및 관련 연구 동향

빅데이터 분석을 위한 기계학습은 크게 지도학습 및 비지도학습의 두 가지로 나눌 수 있다. 지도학습은 컴퓨터의 학습을 위해 정답이 있는 데이터를 제공하는 방식으로 대표적으로 일반화 선형모델, 의사결정나무, 랜덤 포레스트가 있다. 비지

도학습은 정답이 없는 데이터를 기반으로 컴퓨터를 학습시키는 방식으로 대표적으로 군집분석 및 주성분 분석 등이 있다.

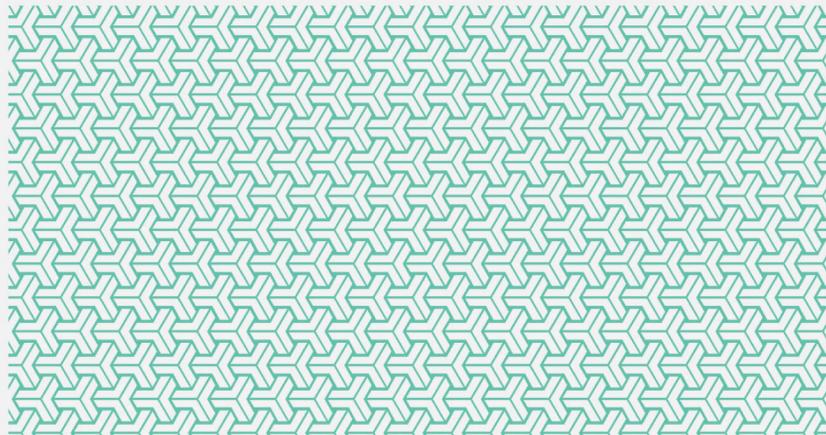
의사결정나무는 분류와 회귀분석이 모두 사용가능한 모델로 여러 가지 규칙을 순차적으로 적용하는 방식으로 데이터를 처리한다. 다른 모델에 비해 상대적으로 단순한 방식으로 정답을 추론한다는 장점이 있으나 오버피팅 가능성이 높다는 단점이 있다. 이를 보완하는 방식으로 랜덤 포레스트 모델이 있다. 랜덤 포레스트는 부트스트랩 샘플링 과정을 통해 학습용 데이터셋보다 작은 크기의 샘플을 여러개 추출하고, 각 샘플에 맞는 의사결정모형을 구축한다. 다수의 의사결정모형에 의해 결과값이 도출되기 때문에 의사결정나무의 단점을 최소화하는 방식으로 사용되고 있다. 이외에도 서포트 벡터머신은 분류를 위한 기준 선인 결정 경계를 정의하는 모델로 데이터 분류 결과가 정확하여 이진데이터 분류에 강점을 나타내고 있으며, 인공신경망은 인간의 뉴런 구조에 착안하여 만든 기계학습 모형으로 딥러닝의 기본적 모델이 된다. 딥러닝은 깊은 구조를 구축한 심층 신경망 기반의 기계학습 방법으로 방대한 규모의 데이터 수집이 필요하며, 이를 위한 데이터 큐레이션이 중요하다.

빅데이터 기반 재난 및 화재 예측 연구동향을 살펴보면 예측을 위해 활용되어야 하는 데이터 유형 및 목록과 기존 데이터를 활용하는 빅데이터 연구방법론에 초점이 맞춰져 있다. 데이터의 유형, 구축정도, 재난·재해 데이터 외 활용가능한 데이터가 무엇이며, 예측을 위한 시나리오 구축 연구에 중점을 두고 있다.

빅데이터 분석 방법론 및 관련 연구 동향 고찰을 통한 시사점은 다음과 같다. 기계학습은 방대한 데이터를 연구 및 정책적 목적에 맞게 분석하는 도구로서 활용되고 있다. 이에 우선적으로 기계학습을 통해 도출하고자 하는 결과를 분명히 하고, 이에 맞는 빅데이터의 선별과 컴퓨터가 기계학습을 원활히 수행할 수 있도록 선별된 데이터의 정제가 필요하다. 또한 화재위험지역 예측모델 구축을 위해 기계학습을 위한 시나리오를 정립해야 한다. 정립된 시나리오는 기계학습을 위해 투입되는 데이터의 형태, 순서, 변수의 유형을 결정하기 때문이다. 마지막으로 시나리오에 따른 데이터가 결정되면 다양한 기계학습 모델 중 예측모델 구축을 위한 최상의 모델을 선택하여야 한다. 데이터의 특성과 도출하고자 하는 목적에 따라 적용할 수 있는 모델의 유형이 달라지기 때문에, 이를 고려하여 적합한 기계학습 모델을 선정하는 것이 중요하다.

제3장 건축물 화재 예측 모델 개발

Chapter.3



1. 대상지 선정 및 데이터 연계
2. 건축물 화재 위험지역 예측 방법론 수립
3. 소결

1. 대상지 선정 및 데이터 연계

1) 대상지 선정 및 특성

본 연구에서 제안하는 위험지역 예측 모델은 대부분 도시지역에 적용 가능한 범용성을 갖추고 있어야 하며, 따라서 대상지는 일반적인 도시가 가지고 있는 범용성을 대부분 갖춘 곳을 선정할 필요가 있다.

도시의 범용성을 적용하기 위해 본 연구에서 공간적 범위로 지정하고자 하는 도시는 다양한 인문사회학적 특성과 물리적 특성을 가지고 있어야 한다. 대상지의 인문사회학적 특성은 일정규모 이상의 인구가 거주 및 생활하고 있으며, 다양한 가구 유형이 혼재하고 있고, 해당 가구 유형의 규모가 유의미할 정도가 되어야 한다는 것이다.

범용성을 적용하기 위한 물리적 특성은 우선 본 연구의 주요 분석 대상인 건축물의 용도, 규모, 형태, 구조, 노후도 등 다양한 속성을 가지고 있는 건축물이 다수 분포되어 있어야 한다. 또한 공간적 특성이 명확히 구분되도록 주거·상업·공업 밀집지역을 가지고 있어야 하며, 개별 밀집지역이 일정한 규모 이상을 갖추고 있어야 한다.

이러한 전제하에 범용성을 가지고 있는 도시는 오랜 역사를 거쳐 개발이 이루어지고, 급격한 경제성장시기에 계획적으로 개발된 공간이 분포하고 있으며, 지속적으로 변화하는 도시여야 한다. 이는 다양한 공간적·물리적 특성을 내포하고 있다는 의미이기 때문이다.

마지막으로 본 연구에서 분석하고자 하는 안전사고 빅데이터는 화재 데이터라는 것이다. 이는 일정수준 이상의 화재가 발생한 지역이어야 한다는 것을 의미한다. 통계분석 및 머신러닝 등 빅데이터 분석을 수행함에 있어 너무 적은 화재 발생 건수는 유의성을 확보할 수 없기 때문이다.

이에 본 연구에서는 다양한 유형 및 일정규모 이상의 인구와 가구로 구성되어 있고, 노후화된 건축물부터 최근에 준공된 건축물까지 혼재되어 있으며, 다양한 용도지역 및 주거·상업·공업 밀집지역을 가지고 있고, 일정규모 이상의 화재 발생 건수가 있는 서울특별시를 대상지로 선정하였다.

대상지 선정 기준		
인문사회학적 특성	물리적 특성	화재 발생
<ul style="list-style-type: none"> • 일정규모의 인구 수 • 다양한 가구 유형 • 일정규모의 유형별 가구 수 	<ul style="list-style-type: none"> • 다양한 건축물 속성 • 건축물 다수 분포 • 용도지역별 밀집지역 • 일정규모의 밀집지역 	<ul style="list-style-type: none"> • 일정규모 화재발생 • 다양한 화재발생 유형

[그림 3-1] 대상지 선정기준

출처 : 연구진 작성

① 서울시 인구 및 가구 현황

서울시 인구는 2022년 1분기 기준 총 9,732,617명으로 대한민국 전체 인구의 18.9%를 차지하고 있으며, 남성은 4,716,704명, 여성은 5,015,913명이 거주하고 있다.¹⁴⁾ 다만, 일일평균 생활인구는 약 11,227천명¹⁵⁾으로 주민등록 인구에 비해 높은 것으로 나타났다. 또한 2021년 기준 서울시 인구밀도는 15,699천명/km² (전국 평균 515명/km²)으로, 인구수가 가장 많은 경기도(1,335명/km²) 및 두 번째로 인구밀도가 높은 부산광역시(4,320명/km²)에 비해 매우 높은 인구밀도 수치를 보이고 있다.¹⁶⁾

서울시 총 일반가구수는 3,982,290 가구로 평균 가구원수는 2.3명 이다. 특히 1인가구는 1,390,701로 전체 가구의 34.9%를 차지하고 있으며, 4인 이상 가구는 19.2%를 차지하고 있다.¹⁷⁾ 이외 2인가구 1,033,901가구, 3인가구 792,690가구로 다양한 가구유형을 보이고 있는 것으로 나타났다.¹⁸⁾

② 서울시 건축물 현황

서울시 건축물 현황을 살펴보면 2021년 기준 총 585,636동이 있다. 동 수로는 주거용 건축물이 74.4%로 가장 높은 비율을 차지하고 있으며, 연면적은 1백

14) 서울 열린데이터 광장, <https://data.seoul.go.kr/dataList/419/S/2/datasetView.do> (검색일 : 2022.06.22.)

15) 서울 열린데이터 광장, <https://data.seoul.go.kr/dataVisual/seoul/seoulLivingPopulation.do> (검색일 : 2022.06.22.)

16) e-나라지표, https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1007 (검색일 : 2022.06.22.)

17) 서울 열린데이터 광장, <https://data.seoul.go.kr/dataList/10996/S/2/datasetView.do> (검색일 : 2022.06.22.)

18) 서울 열린데이터 광장, <https://data.seoul.go.kr/dataList/10996/S/2/datasetView.do> (검색일 2022.06.22.)

m²~2백m²미만이 20.78% 가장 높고, 층수별로는 2~4층이 61%로 가장 높은 비율을 차지하고 있다. 특이점은 대도시이기 때문에 21층 이상 건축물이 4,446동으로 서울시 전체 건축물의 0.76%를 차지하고 있으나, 국내 21층 이상 건축물의 14.6%(전국평균 1.7%)를 차지하고 있어, 고층건축물 비율이 매우 높은 것으로 나타났다. 다만, 2020년 기준 서울시 건축물 중 준공 후 30년 이상 건축물은 총 295,163동으로 전체의 50.4%를 차지하고 있고, 공업용(58.2%) 및 상업용(55.2%) 건축물의 경우 30년 이상 노후화된 건축물의 50%를 넘는 것으로 나타났다.¹⁹⁾

[표 3-1] 서울특별시 건축물 용도·면적·층수별 현황

구분	동수	비율
총 건축물 동수	585,636	100%
용도별	주거용	74.40%
	상업용	21.67%
	공업용	0.42%
	문교사회용	2.78%
	기타	0.73%
면적별	1백m ² 미만	18.22%
	1백m ² ~2백m ² 미만	20.78%
	2백m ² ~3백m ² 미만	15.96%
	3백m ² ~5백m ² 미만	17.43%
	5백m ² ~1천m ² 미만	15.91%
	1천m ² ~3천m ² 미만	6.09%
	3천m ² ~1만m ² 미만	3.76%
층수별	1만m ² 이상	1.85%
	1층	17.99%
	2-4층	61.00%
	5층	10.94%
	6-10층	6.08%
	11-20층	2.93%
	21-30층	0.67%
	31층이상	0.09%
그 외 기타	0.14%	

출처 : 국가통계포털, (검색일 : 2022.06.22.)

용도별 건축물 현황, https://kosis.kr/statHtml/statHtml.do?orgId=116&tblId=DT_MLTM_522

면적별 건축물 현황, https://kosis.kr/statHtml/statHtml.do?orgId=116&tblId=DT_MLTM_540

층수별 건축물 현황, https://kosis.kr/statHtml/statHtml.do?orgId=116&tblId=DT_MLTM_524

③ 서울시 용도지역 현황

서울시 용도지역 현황을 살펴보면 전체 용도지역 중 주거지역의 면적비율이

19) 국토교통부 보도자료(2021.03.09.), 전국 건축물 총 7,275,266동/39억 6천만m², 붙임자료 20년 기준 건축물 현황 통계(건축정책과) 참조.

53.84%로 가장 높은 비율을 차지하고 있으며, 다음으로 녹지지역이 38.64%를 차지하고 있다. 다만 주거지역 중 상업시설을 설치할 수 있는 일반주거지역의 비중이 매우 높아 주·상 혼재지역이 넓게 분포되어 있음을 알 수 있다.

[표 3-2] 서울시 용도지역별 면적 및 비율

구분		면적(㎡)	비율	
용도지역총합계		605,598,290	100%	
주거지역	소계	326,043,340	53.84%	
	전용주거지역	제1종전용	5,007,957	0.83%
		제2종전용	751,996	0.12%
	일반주거지역	제1종일반	67,403,028	11.13%
		제2종일반	141,298,989	23.33%
		제3종일반	98,206,201	16.22%
	준주거지역	13,375,169	2.21%	
상업지역	소계	25,599,274	4.23%	
	중심	358,676	0.06%	
	일반	22,873,609	3.78%	
	근린	827,881	0.14%	
	유통	1,539,108	0.25%	
공업지역	소계	19,974,140	3.30%	
	준공업	19,974,140	3.30%	
녹지지역	소계	233,981,536	38.64%	
	보전	72,603	0.01%	
	생산	1,057,622	0.17%	
	자연	232,851,311	38.45%	

출처 : 서울 열린데이터광장, <https://data.seoul.go.kr/dataList/244/S/2/datasetView.do>
(검색일 : 2022.06.22.)

④ 서울시 화재 발생 현황

서울시 화재 발생 현황 조사는 서울 열린데이터 광장에서 제공하는 각각의 통계 중 최근 5년치(2017~2021년)를 조사하였다. 우선 5년간 서울시 화재발생 건수는 총 28,266건으로 연평균 5653.2건이 발생하였다. 발화요인을 살펴보면 부주의가 57.4%로 가장 높은 수치를 보였으며, 건축물의 노후화로 인한 설비 노후화 및 작동이상으로 추정되는 전기적 요인 및 기계적 요인으로 인한 화재는 8,304건으로 29.38%를 차지하였다.

[표 3-3] 2017-2021년 서울시 발화요인별 화재발생 건수

구분		2017	2018	2019	2020	2021	합계	비율
실화	전기적요인	1,315	1,525	1,322	1,265	1,362	6,789	24.02%
	기계적요인	364	379	283	241	248	1,515	5.36%
	화학적요인	39	43	43	36	47	208	0.74%
	가스누출	23	27	22	14	22	108	0.38%
	교통사고	29	25	37	18	26	135	0.48%
	부주의	3,555	3,639	3,502	2,890	2,638	16,224	57.40%
	기타	21	24	42	59	56	202	0.71%
자연적인요인	10	12	4	5	8	39	0.14%	
방화(명확/의심)	123	167	136	98	87	611	2.16%	
발화요인(미상)	499	527	490	462	457	2,435	8.61%	
합계		5,978	6,368	5,881	5,088	4,951	28,266	100%

출처 : 서울 열린데이터 광장, <https://data.seoul.go.kr/dataList/10167/S/2/datasetView.do> (검색일 : 2022.06.22.)

서울시 장소별 화재발생 건수를 살펴보면 주거에서 발생한 화재가 11,717건으로 41.15%로 나타났으며, 주거 28,266이의 건축물에서 발생한 화재 중 음식점에서 발생한 화재는 3,036건으로 10.74%의 비율로 나타났다.

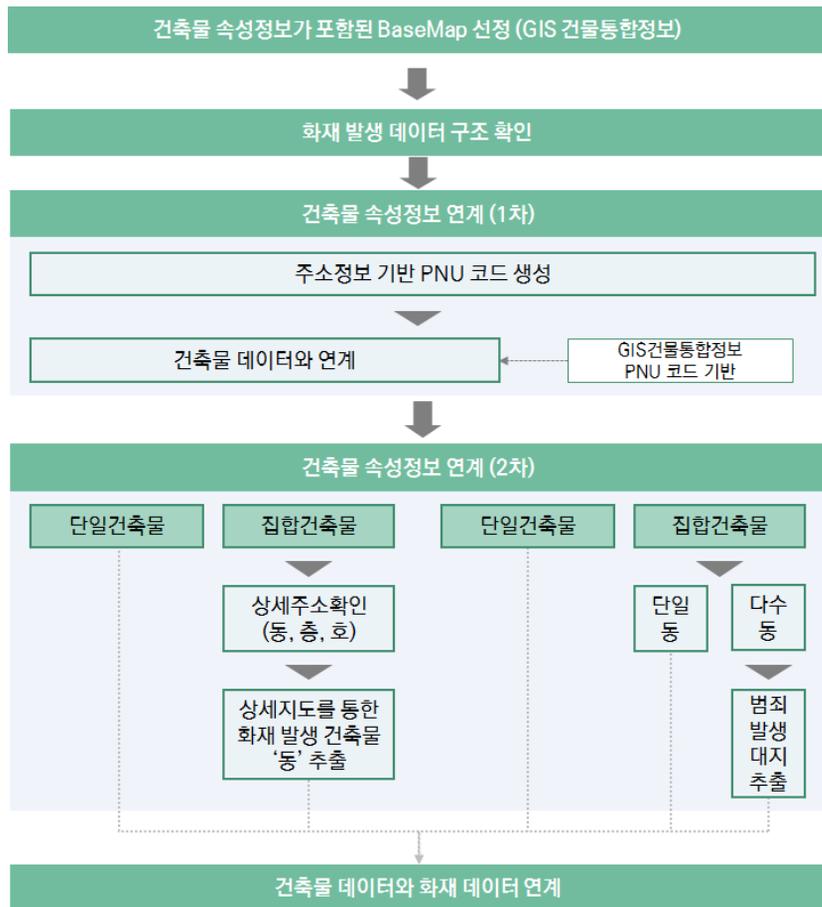
[표 3-4] 2017-2021년 서울시 장소별 화재발생 건수

구분		2017	2018	2019	2020	2021	합계	비율
주거	단독주택	1,016	1,073	983	815	779	4,666	16.51%
	공동주택	1,520	1,669	1,409	1,177	1,137	6,912	24.45%
	기타주택	21	22	30	40	26	139	0.49%
비주거	학교	39	30	32	35	41	177	0.63%
	일반업무	411	438	383	329	372	1,933	6.84%
	판매시설	160	207	205	205	207	984	3.48%
	숙박시설	33	63	58	34	45	233	0.82%
	종교시설	39	48	34	22	32	175	0.62%
	의료시설	45	50	40	29	41	205	0.73%
	공장/창고/직업장	208	185	209	151	162	915	3.24%
	위락오락시설	51	65	46	58	45	265	0.94%
	음식점	681	682	608	571	494	3,036	10.74%
	일반서비스시설	255	294	271	176	194	1,190	4.21%
	기타	279	261	259	201	214	1,214	4.29%
운송(차량철도 등)	468	416	394	338	382	1,998	7.07%	
임야/기타	752	865	920	907	780	4,224	14.94%	
합계		5,978	6,368	5,881	5,088	4,951	28,266	100%

출처 : 서울 열린데이터 광장, <https://data.seoul.go.kr/dataList/318/S/2/datasetView.do> (검색일 : 2022.06.22.)

2) 데이터 연계 및 공간정보화 방향

화재 위험이 높은 건축물을 예측하기 위해서는 건축물 속성정보 데이터와 화재 데이터를 연계하여야 한다. 연계 시 중요한 점은 구축되어 있는 화재 데이터를 건축물 단위 중 어떠한 단위에 연계시킬 수 있는지에 대한 것이다. 화재 위험 건축물 및 지역예측을 가장 정밀하게하기 위해서는 사고가 발생한 건축물 '호'와 연계시키는 것이다. 다만 사고 데이터의 제공 정보 수준과 건축물 속성 데이터의 정보수준을 분석한 후, 최대한 작은 단위에서 연계시킬 수밖에 없다.



[그림 3-2] 건축물 데이터와 화재 데이터 연계 절차

출처 : 연구진 작성

① 건축물 대장 데이터(건축행정시스템)

본 연구에서는 건축물 속성을 포함하고 있는 건축물 데이터로 국토교통부에서 운영하는 건축행정시스템(세움터)에서 제공하는 건축물 대장 데이터를 활용하였다.

본 연구에서 건축물 속성 정보 중 안전사고 데이터와 연관성을 가지고 있을 것으로 예측하고 있는 정보는 건축물 용도, 연면적, 높이, 층수, 노후도, 건폐율 등이다. 건축물 대장 데이터는 이러한 건축물 속성 데이터가 구축되어 있으며, 화재에 따른 멸실을 고려하여 각 년도 화재 발생 전년도 4분기 건축물 대장 데이터를 활용하여 분석을 수행하였다.

② 건축물 화재 발생 데이터

본 연구에서 분석의 대상으로 하는 서울시 건축물 화재 발생 데이터는 2017~2021년의 5년간 화재 발생 데이터로 총 건수는 21,055건이다. 포함되어 있는 정보는 화재 발생 연도, 일시, 시도, 시군구, 법정동, 상세주소, 발화열원(대분류, 소분류), 발화요인(대분류, 소분류), 화재장소(용도), 발화지점, 소실면적(m²), 인명피해(사망, 부상자 수), 재산피해액이다.

건축물 화재 발생 데이터와 GIS건물통합정보의 연계는 1차적으로 화재 발생 데이터의 주소정보를 기반으로 PNU코드²⁰를 생성하여 연계하고, 2차적으로 집합건축물 및 다수의 동이 하나의 지번에 있는 건축물(학교, 아파트 등)은 상세지도를 통해 화재가 발생한 건축물 ‘동’과 연계한다.

건물번호	연도	시도	시군구	법정동	상세주소	발화열원	발화요인	화재장소	발화지점	소실면적	인명피해	재산피해액
1710104118003	0	0	13081	2017	2017-01-01 0:00	사용불명	불발	주거용	화장실	0.00	0	0
1710104447000	0	0	3711	2017	2017-01-01 9:32	사용불명	불발	주거용	화장실	0.00	0	0
1710101006600	0	0	3000	2017	2017-01-01 9:42	사용불명	불발	주거용	화장실	0.00	0	0
1710101006800	0	0	374	2017	2017-01-01 12:40	사용불명	불발	주거용	화장실	0.00	0	0
1710101744570	0	0	30	2017	2017-01-01 14:30	사용불명	불발	주거용	화장실	0.00	0	0
1710101004800	0	0	30	2017	2017-01-01 15:00	사용불명	불발	주거용	화장실	0.00	0	0
1710101005000	0	0	107	2017	2017-01-01 15:10	사용불명	불발	주거용	화장실	0.00	0	0
1710101005500	0	0	124	2017	2017-01-01 17:10	사용불명	불발	주거용	화장실	0.00	0	0
1710101004600	0	0	414	2017	2017-01-01 18:30	사용불명	불발	주거용	화장실	0.00	0	0
1710101102900	0	0	129	2017	2017-01-02 0:12	사용불명	불발	주거용	화장실	0.00	0	0
1710100000000	0	0	030	2017	2017-01-02 0:12	사용불명	불발	주거용	화장실	0.00	0	0
1710104117100	0	0	49	2017	2017-01-02 0:16	사용불명	불발	주거용	화장실	0.00	0	0
1710104007000	0	0	1901	2017	2017-01-02 11:48	사용불명	불발	주거용	화장실	0.00	0	0
1710101007010	0	0	275	2017	2017-01-02 12:37	사용불명	불발	주거용	화장실	0.00	0	0
1710101007600	0	0	113	2017	2017-01-02 14:30	사용불명	불발	주거용	화장실	0.00	0	0
1710101008300	0	0	41	2017	2017-01-02 15:11	사용불명	불발	주거용	화장실	0.00	0	0

[그림 3-3] 건축물 화재 발생 데이터 제공 정보 일부 예시

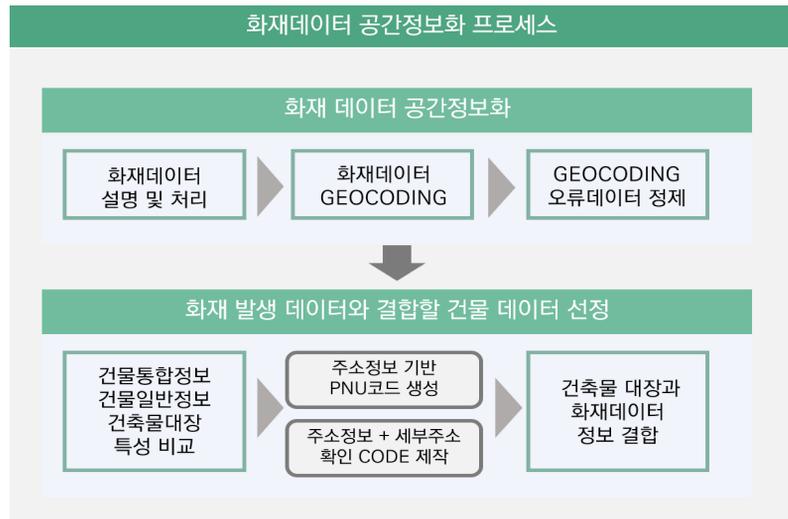
출처 : 서울시 내부자료

20) 총 19자리의 숫자로 구성된 코드로 00(시·도)000(시·군·구)000(읍·면·동)00(리)0(대장구분)0000(본번)0000(부번)으로 이루어져 있다. 앞의 8자리는 행정안전부에서 제공하는 주민등록 주소 코드로 PNU코드 생성 이전 변경 사항을 확인하여야 한다.

3) 건축물 화재 발생 데이터 공간정보화

① 건축물 화재 발생 데이터 구축 프로세스

화재 발생 데이터의 공간정보화 및 건축물 대장과의 연계 구축 프로세스는 크게 3단계로 진행되었다. 우선 건축물 단위에서 화재데이터를 공간정보화 하는 단계를 수행하고, 화재 데이터와 결합할 건축물 정보 데이터를 정리 및 선정한다. 마지막으로 공간정보화 된 화재데이터와 건축물 정보를 결합하여 하나의 통합 데이터를 완성하는 단계로 진행된다.

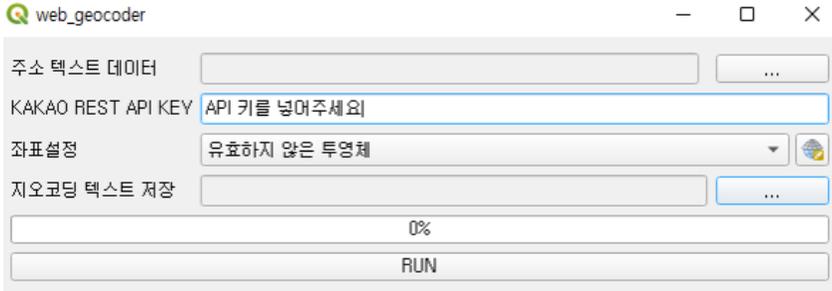


[그림 3-4] 건축물 화재 발생 데이터 구축 프레임워크

출처 : 연구진 작성

② 화재 데이터 지오코딩(Geo-Coding) 및 오류데이터 정제

화재데이터의 공간정보화를 위해 주소정보를 활용하여 지오코딩을 위한 전처리를 진행하였다. 화재 데이터 속성정보 중 주소정보를 활용하였으며, 시도/시군구/읍면동/리/세부주소를 활용하여 새로운 칼럼을 만들어 주소의 정보만 따로 구축하였다. 지오코딩은 QGIS 플러그인 프로그램인 web_geocoder를 활용하였고 KAKAO REST API KEY를 발급받아 지오코딩을 수행하였으며, 좌표계는 EPSG:5179 - Korea 2000 / Unified CS 값을 적용하였다.



[그림 3-5] QGIS Plug-in web_geocoder 활용

출처 : QGIS 프로그램 중 플러그인 실행 화면

지오코딩 된 화재데이터를 확인한 결과 총 21,055개 의 데이터 중 4,370개에서 지오코딩 오류를 확인하였다. 건축물의 위치를 기반으로 지오코딩 된 포인트들의 위치 오류를 정제하기 위해 건물통합정보 데이터를 베이스맵으로 활용하여 확인할 결과 크게 네가지 유형으로 구분가능한 오류를 확인하였다. 오류 칼럼을 제작하여 오류에 따라 다른 이름으로 정의하여 오류를 구분하였다.

[표 3-5] 화재데이터 지오코딩에 따른 오류구분 및 설명

오류구분	설명
noaddr	화재 데이터의 주소 지번이 끝까지 나오지 않거나, 주소가 정확하지 않아서 지오코딩이 되었지만 다른 곳에 표시된 경우
fail	도로 위 혹은 공사장 등 주소가 있지만 지오코딩 오류로 포인트가 생성되지 않은 경우
noft	주소는 정확하여 포인트는 본 위치에 있으나 건물통합정보 데이터의 건물 shp이 없어서 건물 위에 포인트가 없는 경우
noftx	주소도 불분명하고, 건물통합정보 데이터의 건물 shp도 없는 경우

출처 : 연구진 작성

[표 3-6] 화재데이터 지오코딩에 따른 오류구분 및 예시

오류구분	오류예시																																																																																																																														
noaddr																																																																																																																															
	<table border="1"> <tbody> <tr><td>519</td><td>1000000</td><td>서울특별시 강남구 삼성동 1048-48</td><td>847106.16442</td><td>1244708.5417</td><td></td></tr> <tr><td>522</td><td>1000000</td><td>서울특별시 관악구 봉곡동 139-7</td><td>851151.86288</td><td>1246518.8462</td><td></td></tr> <tr><td>521</td><td>1000000</td><td>서울특별시 강남구 삼성동 540</td><td>846751.85284</td><td>1244191.5988</td><td></td></tr> <tr><td>523</td><td>1000000</td><td>서울특별시 관악구 봉곡동 485-18</td><td>846485.69266</td><td>1252332.2521</td><td></td></tr> <tr><td>520</td><td>1000000</td><td>서울특별시 유곡동 유곡동 718-89</td><td>841821.54177</td><td>1244233.2096</td><td></td></tr> <tr><td>524</td><td>1000000</td><td>서울특별시 강서구 가동동 381-27</td><td>842626.88825</td><td>1250070.0727</td><td></td></tr> <tr><td>525</td><td>1000000</td><td>서울특별시 강남구 삼성동 799-7</td><td>851877.25627</td><td>1244442.8506</td><td></td></tr> <tr><td>526</td><td>1000000</td><td>서울특별시 영등포구 영등포동7가 84-153</td><td>841422.88381</td><td>1241768.2484</td><td></td></tr> <tr><td>527</td><td>noaddr</td><td>서울특별시 강서구 가동동 718-89</td><td>841622.01540</td><td>1244015.2314</td><td></td></tr> <tr><td>528</td><td>1000000</td><td>서울특별시 양천구 양천동 171-8</td><td>860061.02744</td><td>1252321.4495</td><td></td></tr> <tr><td>529</td><td>1000000</td><td>서울특별시 강남구 가동동 497-9</td><td>851978.10710</td><td>1243532.8345</td><td></td></tr> <tr><td>530</td><td>noaddr</td><td>서울특별시 용마루동 용마루동 779</td><td>851893.25477</td><td>1250264.3098</td><td></td></tr> <tr><td>531</td><td>noaddr</td><td>서울특별시 강남구 삼성동 1159</td><td>841591.59516</td><td>1250207.8504</td><td></td></tr> <tr><td>532</td><td>1000000</td><td>서울특별시 강남구 가동동 247-9</td><td>841655.75581</td><td>1250248.7385</td><td></td></tr> <tr><td>533</td><td>1000000</td><td>서울특별시 강남구 삼성동 445</td><td>847448.28678</td><td>1243718.7142</td><td></td></tr> <tr><td>534</td><td>1000000</td><td>서울특별시 강남구 삼성동 199-27</td><td>851810.18425</td><td>1242437.9358</td><td></td></tr> <tr><td>535</td><td>1000000</td><td>서울특별시 노원구 공릉동 51-1</td><td>861950.98368</td><td>1251804.4381</td><td></td></tr> <tr><td>536</td><td>1000000</td><td>서울특별시 강서구 가동동 441</td><td>841880.58209</td><td>1250004.7688</td><td></td></tr> <tr><td>537</td><td>1000000</td><td>서울특별시 노원구 공릉동 384-120</td><td>862471.25427</td><td>1251821.8779</td><td></td></tr> <tr><td>538</td><td>1000000</td><td>서울특별시 서대문구 남가좌동 291-2</td><td>848421.05900</td><td>1253489.2584</td><td></td></tr> <tr><td>539</td><td>1000000</td><td>서울특별시 강서구 가동동 88</td><td>841418.42584</td><td>1251800.4104</td><td></td></tr> </tbody> </table>	519	1000000	서울특별시 강남구 삼성동 1048-48	847106.16442	1244708.5417		522	1000000	서울특별시 관악구 봉곡동 139-7	851151.86288	1246518.8462		521	1000000	서울특별시 강남구 삼성동 540	846751.85284	1244191.5988		523	1000000	서울특별시 관악구 봉곡동 485-18	846485.69266	1252332.2521		520	1000000	서울특별시 유곡동 유곡동 718-89	841821.54177	1244233.2096		524	1000000	서울특별시 강서구 가동동 381-27	842626.88825	1250070.0727		525	1000000	서울특별시 강남구 삼성동 799-7	851877.25627	1244442.8506		526	1000000	서울특별시 영등포구 영등포동7가 84-153	841422.88381	1241768.2484		527	noaddr	서울특별시 강서구 가동동 718-89	841622.01540	1244015.2314		528	1000000	서울특별시 양천구 양천동 171-8	860061.02744	1252321.4495		529	1000000	서울특별시 강남구 가동동 497-9	851978.10710	1243532.8345		530	noaddr	서울특별시 용마루동 용마루동 779	851893.25477	1250264.3098		531	noaddr	서울특별시 강남구 삼성동 1159	841591.59516	1250207.8504		532	1000000	서울특별시 강남구 가동동 247-9	841655.75581	1250248.7385		533	1000000	서울특별시 강남구 삼성동 445	847448.28678	1243718.7142		534	1000000	서울특별시 강남구 삼성동 199-27	851810.18425	1242437.9358		535	1000000	서울특별시 노원구 공릉동 51-1	861950.98368	1251804.4381		536	1000000	서울특별시 강서구 가동동 441	841880.58209	1250004.7688		537	1000000	서울특별시 노원구 공릉동 384-120	862471.25427	1251821.8779		538	1000000	서울특별시 서대문구 남가좌동 291-2	848421.05900	1253489.2584		539	1000000	서울특별시 강서구 가동동 88	841418.42584	1251800.4104	
	519	1000000	서울특별시 강남구 삼성동 1048-48	847106.16442	1244708.5417																																																																																																																										
	522	1000000	서울특별시 관악구 봉곡동 139-7	851151.86288	1246518.8462																																																																																																																										
	521	1000000	서울특별시 강남구 삼성동 540	846751.85284	1244191.5988																																																																																																																										
	523	1000000	서울특별시 관악구 봉곡동 485-18	846485.69266	1252332.2521																																																																																																																										
	520	1000000	서울특별시 유곡동 유곡동 718-89	841821.54177	1244233.2096																																																																																																																										
	524	1000000	서울특별시 강서구 가동동 381-27	842626.88825	1250070.0727																																																																																																																										
	525	1000000	서울특별시 강남구 삼성동 799-7	851877.25627	1244442.8506																																																																																																																										
	526	1000000	서울특별시 영등포구 영등포동7가 84-153	841422.88381	1241768.2484																																																																																																																										
	527	noaddr	서울특별시 강서구 가동동 718-89	841622.01540	1244015.2314																																																																																																																										
	528	1000000	서울특별시 양천구 양천동 171-8	860061.02744	1252321.4495																																																																																																																										
	529	1000000	서울특별시 강남구 가동동 497-9	851978.10710	1243532.8345																																																																																																																										
	530	noaddr	서울특별시 용마루동 용마루동 779	851893.25477	1250264.3098																																																																																																																										
	531	noaddr	서울특별시 강남구 삼성동 1159	841591.59516	1250207.8504																																																																																																																										
	532	1000000	서울특별시 강남구 가동동 247-9	841655.75581	1250248.7385																																																																																																																										
	533	1000000	서울특별시 강남구 삼성동 445	847448.28678	1243718.7142																																																																																																																										
	534	1000000	서울특별시 강남구 삼성동 199-27	851810.18425	1242437.9358																																																																																																																										
	535	1000000	서울특별시 노원구 공릉동 51-1	861950.98368	1251804.4381																																																																																																																										
	536	1000000	서울특별시 강서구 가동동 441	841880.58209	1250004.7688																																																																																																																										
537	1000000	서울특별시 노원구 공릉동 384-120	862471.25427	1251821.8779																																																																																																																											
538	1000000	서울특별시 서대문구 남가좌동 291-2	848421.05900	1253489.2584																																																																																																																											
539	1000000	서울특별시 강서구 가동동 88	841418.42584	1251800.4104																																																																																																																											

fail

550	fail(주소없음)	서울특별시 서초구 원지동 520-2	NULL	NULL
-----	------------	---------------------	------	------

noft

474	noaddr	서울특별시 강서구 신정동 905-9	94357140602	1947628.0231
475	noaddr	서울특별시 송파구 난곡동 89-33	94201483951	1954982.2725
476	noaddr	서울특별시 강서구 신정동 328-25	94948932768	1950582.0286
477	noaddr	서울특별시 송파구 송정동 371-583-25	94908131428	1948789.8811
478	noaddr	서울특별시 송파구 송정동 378-2	94202776568	1942377.4966
479	noaddr	서울특별시 강서구 유곡동 1907-21	94198453261	1944296.8955
480	noaddr	서울특별시 노원구 상계동 456-112	94178210714	1942682.0259
481	noaddr	서울특별시 강서구 고척동 50-8	94355705964	1941448.4182
482	noaddr	서울특별시 강서구 신도림동 410-15	94320422275	1942079.4664
483	noaddr	서울특별시 서대문구 양정동 510-2	94942376259	1951206.0347
484	noaddr	서울특별시 송파구 개천동 209-12	94388143353	1944979.0820
485	noaddr	서울특별시 송파구 반곡동 371-407-892	95227836470	1947348.4326
486	noaddr	서울특별시 송파구 문정동 86-7	94669346827	1942639.9970
487	noaddr	서울특별시 송파구 잠실동 1-1	94908144283	1946838.1493
488	noaddr	서울특별시 서대문구 홍제동 214-1	95111488203	1954519.9426
489	noaddr	서울특별시 강서구 가락동 826-37	95498839468	1944119.5288
490	noaddr	서울특별시 강서구 서교동 476-14	94802756001	1950797.4956
491	fail	서울특별시 송파구 신도림동 850	94719110268	1951544.4143
492	noaddr	서울특별시 관악구 신도림동 1518-17	95021841851	1941281.8054
493	noaddr	서울특별시 송파구 가락동 124-22	94668632002	1944884.2526
494	noaddr	서울특별시 노원구 공릉동 745	94206279939	1958810.8222
495	noaddr	서울특별시 송파구 신도림동 484	95264257678	1948723.8714
496	noaddr	서울특별시 강서구 고척동 378-1386	94949839938	1944988.9896

noftx

550	fail(주소없음)	서울특별시 서초구 원지동 520-2	94949839938	1952718.5204
551	noaddr	서울특별시 강서구 신정동 905-9	94357140602	1947628.0231
552	noaddr	서울특별시 송파구 난곡동 89-33	94201483951	1954982.2725
553	noaddr	서울특별시 강서구 신정동 328-25	94948932768	1950582.0286
554	noaddr	서울특별시 송파구 송정동 371-583-25	94908131428	1948789.8811
555	noaddr	서울특별시 송파구 송정동 378-2	94202776568	1942377.4966
556	noaddr	서울특별시 강서구 유곡동 1907-21	94198453261	1944296.8955
557	noaddr	서울특별시 노원구 상계동 456-112	94178210714	1942682.0259
558	noaddr	서울특별시 강서구 고척동 50-8	94355705964	1941448.4182
559	noaddr	서울특별시 강서구 신도림동 410-15	94320422275	1942079.4664
560	noaddr	서울특별시 서대문구 양정동 510-2	94942376259	1951206.0347
561	noaddr	서울특별시 송파구 개천동 209-12	94388143353	1944979.0820
562	noaddr	서울특별시 송파구 반곡동 371-407-892	95227836470	1947348.4326
563	noaddr	서울특별시 송파구 문정동 86-7	94669346827	1942639.9970
564	noaddr	서울특별시 송파구 잠실동 1-1	94908144283	1946838.1493
565	noaddr	서울특별시 서대문구 홍제동 214-1	95111488203	1954519.9426
566	noaddr	서울특별시 강서구 가락동 826-37	95498839468	1944119.5288
567	noaddr	서울특별시 강서구 서교동 476-14	94802756001	1950797.4956
568	noaddr	서울특별시 송파구 신도림동 850	94719110268	1951544.4143
569	noaddr	서울특별시 관악구 신도림동 1518-17	95021841851	1941281.8054
570	noaddr	서울특별시 송파구 가락동 124-22	94668632002	1944884.2526
571	noaddr	서울특별시 노원구 공릉동 745	94206279939	1958810.8222
572	noaddr	서울특별시 송파구 신도림동 484	95264257678	1948723.8714
573	noaddr	서울특별시 강서구 고척동 378-1386	94949839938	1944988.9896

출처 : QGIS 프로그램 활용 연구진 제작성, 네이버맵(<https://map.naver.com>)

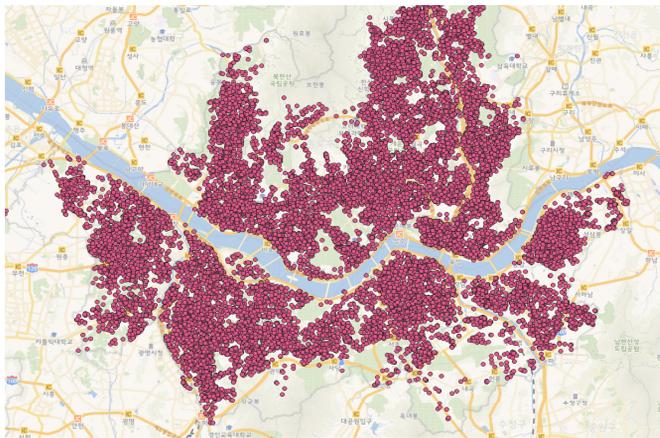
오류 데이터 확인 및 식별 후 오류 포인트를 직접 확인하여 포인트의 위치를 정확한 건축물 등으로 옮겨주는 Point relocation 작업을 수행하였다. noft의 경우는 베이스맵인 건물통합정보에는 건물 shp이 없지만 지도를 기반으로 그 자리에 포인트를 옮겨주었으며, noaddr나 noftX에서 주소가 불분명하여 도저히 찾을 수 없는 값들은 오류 칼럼 그대로 작성하였다.

[표 3-7] 화재데이터 Point Relocation 칼럼코드 및 설명

칼럼구분	설명
success	지오코딩이 잘되어서 건물에 매칭이 성공한 경우
r_success	오류들 중 일일이 확인하였을 때 정확한 주소를 추정 가능하거나, 위치를 추정 가능한 점들은 그 지점 건물 위에 포인트를 매칭
s_change	fail과 같이 포인트 생성이 되지 않은 데이터들은 주소 확인 후 일일이 포인트를 만들어 해당 위치에 매칭
changeX	fail 데이터 확인 후에도 매칭할 수 없는 경우

출처 : 연구진 작성

약 93% 정도의 화재 데이터를 건축물 단위의 공간정보 데이터 구축에 성공하였으며, 본 데이터를 활용하여 이후 건축물 데이터와의 결합을 진행하였다. [표 3-8]은 지오코딩 결과와 포인트 재배치 후 결과를 나타내고 있다. [그림 3-6]은 완성된 건물기반의 화재데이터 공간정보 데이터이며, 정확한 위치로 이동된 포인트들을 기반으로 'QGIS - X/Y 필드를 레이어에 추가' 공간처리 기능을 활용하여 해당 포인트의 좌표 값을 갱신하였다.



[그림 3-6] 화재 데이터의 공간 정보화 완료

출처 : QGIS를 활용한 연구진 작성

③ 화재 데이터와 건축물 속성정보 데이터 연계 구축

화재 데이터와 연계할 건축물 속성정보는 건축행정시스템의 건축물 대장자료를 활용하였다. 건축물 대장데이터와의 연계를 위해 1차적으로 화재 발생 데이터의 주소정보를 기반으로 PNU코드를 생성하였다. PNU 코드는 필지 고유번호로써, 행정구역 10자리와 지번 9자리, 총 19자리로 구성되어 있다. 세분화 해보면 행정 구역 10자리는 시도(2자리), 시군구(3자리), 읍면동(3자리), 리(2자리)를 나타내고, 지번 9자리는 필지 구분(1자리)과 본번(4자리)과 부번(4자리)로 구성되어 있

다. 따라서 화재 데이터에 포함되어 있는 시도, 시군구, 읍면동, 리, 세부주소에 포함되어 있는 본번과 부번을 추출하여 PNU코드를 생성하였다. 변환과정에서 화재데이터의 세부주소가 불분명하거나, 정확하지 않은 주소(도로 위, 간이 공사장 등) PNU코드가 제대로 생성되지 않는 데이터의 경우는 PNU매칭 칼럼을 추가하여 noaddr로 분류하였다.

화재 데이터와 건축물 대장 데이터를 연계할 때 주의한 점은 화재로 인한 건축물이 멸실 되었을 경우 2022년 현재의 건축물 대장이 존재하지 않는다는 점이다. 이에 화재데이터와 결합할 건축물 대장은 2016년도 4분기, 2017년도 4분기, 2018년도 4분기, 2019년도 4분기, 2020년도 4분기 총 5년의 데이터를 활용하였다. 화재데이터가 2017년부터 2021년도까지의 데이터이기 때문에, 이에 상응하는 건축물 대장을 활용하여 결합을 진행하였다. 결합 방법은 화재 발생 시기 기준, 전년도 4분기 건축물 대장데이터와 매칭을 진행하는 것이다. 건축물 대장이 존재하지 않는 등의 상황으로 매칭이 되지 않는 경우에만 당해연도 4분기 건축물 데이터와 매칭 하였다. 예를 들어, 2018년 화재 발생 데이터는 2017년 4분기 건축물 데이터와 매칭이 된다.

데이터 처리를 위해 Python pandas 라이브러리를 사용하였다. 매칭에 대한 자세한 프로세싱은 아래와 같이 진행되어진다.

1) 단일건물인 경우, 바로 화재데이터에 건축물대장 데이터 결합 ⇒ 2) 집합건물에서 해당 주소의 건물이 한 동인 경우, 데이터 결합 ⇒ 3) 집합건물/건축물대장의DONG_NM(동이름)이 화재 세부주소와 일치하는 경우 해당 데이터 결합 ⇒ 4) 부수적으로 건축물대장 상 동 이름이 '제00동' 형식, 'A동'을'에이동'으로 기재한 경우 등 의미에 부합하는 해당 동을 찾아 결합 ⇒ 5) 화재데이터 세부주소에서 동 이름을 숫자만 기재한 경우 해당 동이 있는지 찾아서 결합 ⇒ 6) 연결되지 않을시 화재발생 당해연도 건축물 연결 ⇒ 7) 비매칭결과 예러 처리코드

프로세스 진행결과 주소가 일치하는 건물이 건축물대장에 존재하지 않은 경우 'noDJ'로 라벨링 하였으며, 주소가 일치하는 건물이 있으나 세부주소 일치하는 것을 찾지 못한 경우 'Revision'으로 라벨, 직접 검토 후 수정하였다.

화재데이터와 건축물 대장 데이터를 매칭한 결과 총 21,055의 데이터 중 18,624개의 데이터가 화재데이터와 건축물 데이터와 결합이 되었으며, 건축물 대장이 존재하지 않아서 결합되지 않은 NoDJ는 595개, 세부 동 주소 및 위치 주소 등이 건축물 대장과 표기 방법에서 차이가 발생했거나 건축물 대장에는 존재하지 않는 380개의 데이터는 Revision으로 분류하였고, 추후 직접 모든 데이터들을 확

인하여 수정하였다. 그 외에 1,456개의 Fail 데이터는 화재 데이터 자체의 오류로 매칭이 불가하였다.

Revision 항목은 다시 확인해보아야 할 필요가 있기 때문에, 380개의 데이터를 직접 건축물 대장과 화재 데이터의 주소를 확인하여 식별하였다. 그 결과 190개의 데이터가 매칭 가능하였으며, 16개의 데이터가 건축물 대장이 없어서 결합이 되지 않았고, 174개의 데이터가 화재 데이터의 주소 오류 등으로 매칭이 되지 않았다. 새롭게 매칭 가능한 항목은 r_Success로 표기하고, 건축물 대장이 없는 항목은 r_NoDJ, 주소 오류 등으로 매칭 불가 항목은 r_Remove로 표기되었다. 최종 결합에 성공한 데이터는 Success와 r_Success 항목으로 볼 수 있으며 약 86.4%의 매칭 성공률을 보인다.

[표 3-8] 건축물 대장과 화재 데이터 매칭 결과

분류	재확인 분류	개수	퍼센트
Success	Success	18,624	88.5%
NoDJ	NoDJ	595	2.8%
Revision	r_Success	190	0.9%
	r_NoDJ	16	0.07%
	r_Remove	174	0.82%
Fail	필드값 없음	1,456	6.9%
전체 합계		21,055	100%

출처 : 연구진 작성

④ 화재 발생 건축물의 공간 단위 통합데이터 구축

화재 데이터와 건축물 대장 데이터의 매칭이 성공한 데이터셋의 결과는 [표2-4]와 같다. 매칭에 성공한 총 Success 개수는 18,814개로 약 89.4%의 데이터가 성공적으로 정제 및 매칭 되었다. 매칭에 성공하지 못한 Fail의 개수는 2,241개로 약 10.6%의 데이터가 화재 데이터 및 건축물 대장 데이터의 오류로 매칭에 실패하였다.

[표 3-9] 화재 건축물 연계 데이터 최종 매칭 결과

분류	재확인 분류	개수	퍼센트
Success	Success(성공)	18,624	88.5%
	r_Success(재매칭 성공)	190	0.9%
성공 총 합계		18,814	89.4%
Fail	NoDJ(건축물 대장X)	595	2.8%
	r_NoDJ(재매칭 - 건축물 대장X)	16	0.07%
	r_Remove(재매칭 - 주소 오류등)	174	0.82%
	필드값 없음(화재 데이터 오류 등)	1,456	6.9%
실패 총 합계		2,241	10.6%
전체 합계		21,055	100%

출처 : 연구진 작성

최종 데이터셋에 포함된 데이터는 위치 좌표 (x,y)·화재데이터 정보·PNU 코드·건축물대장 정보·건축물 대장 최종 분류로 구성된다. 위치 좌표 x,y는 정확한 주소 기반으론 기존 화재 발생 포인트가 갱신된 좌표이며 분석시에는 마지막 칼럼인 'DJMatchRes'를 통해 최종 매칭 여부를 확인 할 수 있다.

화재 데이터 칼럼은 기존의 화재 데이터 정보인 [조사서 번호, 사망, 부상, 인명피해(명) 소계, 재산피해 소계, 화재 발생 년, 화재 발생 년월일, 시도, 시·군·구, 읍면동, 리, 세부 주소, 발화열원 대분류, 발화열원 소분류, 발화요인 대분류, 연소확대물 대분류, 소실 면적(㎡), 장소 대분류, 발화지점 대분류]의 내용을 포함하고 있다.

[표 3-10] 화재 데이터 및 건축물 대장 매칭 결과 칼럼Code 및 설명

칼럼Code	설명
success	건축물대장과 주소가 잘 매칭된 경우
NoDJ	건축물 대장에 해당 주소 건물이 없는 경우
R_NoDJ	재확인 (건축물 대장 앞뒤 년도) 시에도 건축물 대장에 정보가 없는 경우
R_success	재확인 시 건축물 대장의 정보를 발견한 경우 (예 : 세부주소에서는 'A동' 건축물대장에서는 '에이동'으로 나와 매칭이 안되는 경우 등을 모두 다 확인)
R_remove	noaddr,noftX등 세부 주소가 잘못 되어 있는데 데이터셋에 포함되어 있어서 건축물 대장과 매칭이 어려운 경우로 분석 진행시 삭제하면 되는 칼럼
fail	필드값 없음
noaddr, noftX	주소 이상 혹은 지오코딩이 되지 않은 데이터 등은 건축물 대장과 매칭할 수 없기 때문에 이런 데이터들은 모두 비어 있는 채로 있는 경우

출처 : 연구진 작성

건축물 대장 데이터는 사용승인 일, 대지면적, 건축 면적, 건폐율, 연면적, 용적률, 높이, 지상 층수, 지하 층수, 승용 승강기 수 등의 다양한 건축물 정보를 포함하고 있다. 마지막 칼럼은 건축물 대장 전처리 과정에서 최종 분류 시 사용했던 칼럼이다. 추후 화재 건축물 통합 데이터셋 사용을 위해서는 CM의 건축물 분류 칼럼 중 'success'와 'r_success'를 추출하여 사용하면 정확한 위치 공간 및 건축물 데이터 정보를 함께 활용할 수 있다.

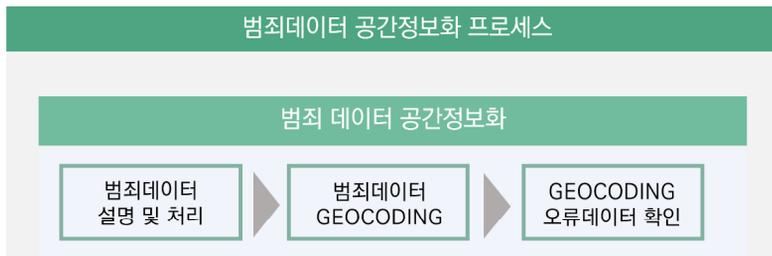
4) 건축물 범죄 발생 데이터 공간정보화

건축물에서 발생하는 안전사고의 가장 많은 수를 차지하는 것은 범죄이기 때문에 화재를 예측하는 모델의 변수로서 범죄 발생을 고려하였다. 이에 주소기반 범

죄발생 데이터를 빅데이터 분석방법의 변수로 활용하기 위해 공간정보화를 진행하였다. 정제된 범죄데이터는 데이터 구득 한계로 인해 2014년~2018년 5년간 데이터를 활용하였다.

① 건축물 범죄 발생 데이터 구축 프로세스

범죄 발생 데이터의 공간정보화는 지오코딩을 통해 진행되었다. 빅데이터 분석 방법론을 적용한 범죄 발생 예측 모형은 범죄발생 지점의 정확한 건축물 위치정보(x,y)를 통해 이루어져야 하지만, 방대한 데이터로 인해 지오코딩을 통해서만 공간정보화를 수행하였다.



[그림 3-7] 건축물범죄 발생 데이터 구축 프레임워크

출처 : 연구진 작성

② 범죄 데이터 지오코딩(Geo-Coding) 및 오류데이터 유형

2014~2018년 서울시 범죄 발생 데이터 총 690,174개 중 누락된 정보가 있는 데이터를 제거한 664,664개의 데이터를 대상으로 지오코딩을 수행하였다. 650,305개인 약 97.8%가 지오코딩에 성공하였고, 14,359개 (약 2.1%)의 지오코딩 오류를 확인하였다.

[표 3-11] 범죄 데이터 지오코딩 성공률

분류	설명	개수	퍼센트
Success	매칭에 성공한 데이터	650,305	97.8%
Fail	매칭에 실패한 데이터 (크게 4유형으로 분류 가능)	14,359	2.2%
전체 합계		664,664	100%

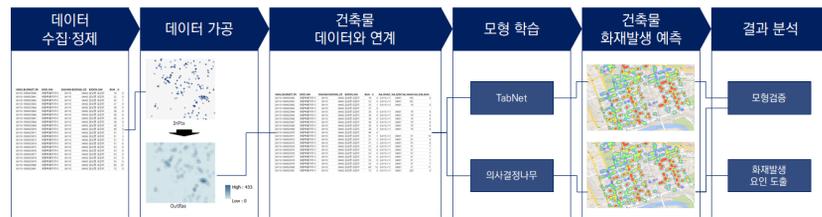
출처 : 연구진 작성

오류 데이터 유형은 크게 4가지 유형으로 분류할 수 있으며, 도로 위 또는 지하 철역사 주변의 주소로 정확한 필지가 없는 경우, 주소에 필지가 없는 경우(아파트 주변 혹은 도로 위), 주소 자체가 없는 경우, 필지 모양이 이상하거나 공원, 축구장, 산 등에 있는 주소로 분류될 수 있다.

2. 건축물 화재 위험지역 예측 방법론 수립

1) 기계학습 기반의 화재 및 범죄 취약건물 예측모형 개발

본 연구에서는 건축물의 재난·재해 취약성 분석을 위해 건축물대장 변수와 화재 및 범죄 변수 등 다양한 관련 변수들을 연계하여 분석하고자 한다. 또한, 공간분석을 활용하여 건축·도시공간의 서로 다른 위계에 따라 분포하는 변수를 공간적으로 연계하여 분석하고자 한다. 예측모형으로는 기계학습 모형 중 의사결정나무와 딥러닝 모형인 TabNet 모형을 사용하여, 기계학습 모형이 데이터에서 학습한 규칙에 대한 해석을 수행하고, 이를 통하여 건축물 재난·재해 취약성과 연관된 요인을 분석하고 예측 방법론을 검증하고자 한다.



[그림 3-8] 기계학습 기반의 화재 취약건물 예측모형 개발 흐름

출처 : 연구진 작성

2) 화재 및 범죄 데이터 가공

□ 건축물대장 정보

본 연구에서는 건축물대장정보를 활용하기 위해 ‘GIS건물통합정보’를 사용하였다. 건물통합정보에는 ‘GIS건물통합정보’와 ‘건물통합정보_마스터’가 있다. 두 자료는 데이터 관리조직이 상이하여 속성정보 및 속성정보 구축 품질의 차이가 있다. 건축물의 용도, 구조, 연면적, 용적률 등 건축물대장 정보 중 일부를 활용하고자하며, 해당 정보들이 구축된 수준을 바탕으로 ‘GIS건물통합정보’를 최종 사용 데이터로 선정하였다. 즉, 분석에 사용하는 ‘GIS건물통합정보’는 개별 건축물의 위치정보와 함께 건축물대장정보를 포함하는 데이터이다.

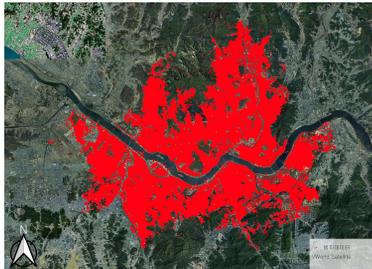
□ 범죄 데이터 가공 및 공간변수 구축

범죄데이터는 2014년부터 2018년까지 발생한 강간·강제추행, 절도, 폭력, 살인, 강도의 5가지 유형의 범죄 데이터로 구성되어있다. 범죄 유형별 발생 건수는 강간·강제추행 31,279건, 강도 1,089건, 살인 783건, 절도 269,645건, 폭력

347,509건 등 총 650,305건으로 구성되어있다. 전체 데이터는 발생위치에 대한 주소정보와 범죄유형정보를 모두 포함하고 있다. 주소정보는 지번주소로 구성되어있어, 공간분석을 위해 지오코딩(Geocoding) 작업을 수행하였다.

지오코딩은 텍스트 형식의 주소정보에 기반하여 공간상 위치좌표를 부여하는 작업으로, 지오코딩을 수행한 주소정보는 X, Y 좌표정보를 갖는 점데이터로 변환 가능하다. 전체 범죄데이터는 지오코딩을 통해 X, Y좌표를 갖도록 변환하였으며, 강간·강제추행, 강도, 살인, 절도, 폭력 5개 유형별로 데이터를 분리하였다. 즉 전체 범죄데이터를 포함한 총 6개 데이터셋을 구축하였으며, 변환된 데이터는 QGIS를 이용하여 Korea 2000 / Unified CS(EPSG: 5179)로 좌표체계를 변환 및 통일하였다.

발생위치	발생장소	범죄수법	전수죄종 7
서울 서초구	유흥주점	강제추행	강간·강제추행
서울 서초구	전담	기타(기타절도)	절도
서울 서초구	전담		폭력
서울 서초구	노상	오토바이절도	절도
서울 금천구	노상		폭력
서울 금천구	기타	기타(기타절도)	절도
서울 금천구	아파트		폭력
서울 서초구	기타		폭력
서울 금천구	기타		폭력
서울 금천구	기타		폭력
서울 금천구	기타	기타(점입절도)	절도
서울 서초구	기타		폭력
서울 금천구	주역	빈집점입절도	절도
서울 금천구	주역	빈집점입절도	절도
서울 금천구	다세대 원룸 연립가옥		폭력



범죄 데이터(주소 일부 가림)

공간화된 범죄발생 점 데이터

[그림 3-9] 범죄데이터 공간화

출처 : 연구진 작성

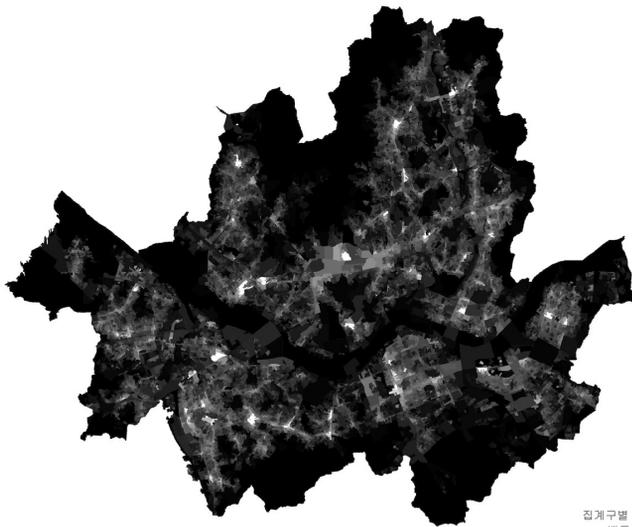
점데이터로 변환한 범죄발생 데이터는 최종적으로 집계구 단위에서 면적당 발생 횟수(발생횟수/㎡)로 계산하고자 하였다. 그러나 지번주소의 경계를 나타내는 ‘연속지적도’와 집계구의 경계를 나타내는 ‘전수집계구 경계’ 자료를 검토한 결과 집계구의 경계선이 지번주소의 경계선과 상이한 부분이 많아, 추가적인 변환 작업이 필요하였다.

이를 위해 지번단위 데이터를 집계구 단위로 변환하는 네 단계 작업을 수행하였다. 첫째로, 지번단위 경계에 포함되는 범죄발생 횟수를 계산하였다. 둘째로, 개별 지번 폴리곤의 면적을 계산하여 지번주소 별 면적당 발생횟수(발생횟수/㎡)를 도출하고 이를 개별 폴리곤에 속성정보로 저장하였다. 셋째로, 각 폴리곤의 지번주소 별 면적당 발생횟수를 셀 값으로 갖는 셀 사이즈 1m의 래스터 자료를 구축하였다. 마지막으로, 집계구 shp 파일을 이용하여 집계구 개별 폴리곤에 해당하는 래스터 자료 셀들의 평균값을 계산함으로써 집계구 단위의 면적당 범죄 발생횟수(발생횟수/㎡)를 도출하였다.



[그림 3-10] 집계구와 지번주소 경계(지번주소: 붉은색, 집계구: 파란색)

출처: 연구진 작성



집계구별 범죄 발생률
밴드 1: crime_zone
0.025086
0

[그림 3-11] 집계구별 범죄 발생률 지도

출처: 연구진 작성

범죄발생률은 m^2 당 평균 0.001회 정도로 발생하고 있으나, 지역에 따라 최대 약 0.07회까지 나타나고 있으며, 이는 km^2 당 약 6만 9천건 정도이다. 평균 발생률이 가장 낮은 살인의 경우 km^2 당 약 1.3건 정도가 발생하고 있으며, 가장 높은 지역의 경우에도 km^2 당 약 272건 정도로 발생하고 있다.

[표 3-12] 범죄 유형별 면적당 범죄발생횟수 기술통계

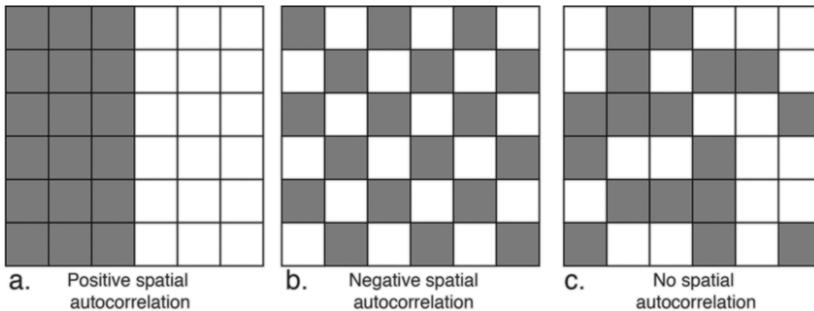
단위: 발생횟수/㎡ (발생횟수/㎢)

구분	최소	최대	평균
범죄 전체	0	0.0698 (68,985)	0.001075 (1,075)
강도	0	0.0006 (617)	0.000002 (1.8)
강간	0	0.0049 (4,904)	0.000051 (51.7)
절도	0	0.0565 (56,479)	0.000446 (446)
폭력	0	0.0324 (32,412)	0.000574 (574)
살인	0	0.0003 (272)	0.000001 (1.3)

출처 : 연구진 작성

□ 화재 데이터 가공 및 공간변수 구축

본 연구는 건축물 화재발생의 공간 자기상관이 있을 것으로 가정하고 변수를 구축하였다. [그림 3-12]의 a는 공간자기상관이 높은 경우로, 화재가 난 건축물 근처에서 화재발생이 높은 경우로 볼 수 있다. 반면 b의 경우는 화재가 난 건축물 근처에서 화재발생률이 낮은 경우로 공간자기상관이 낮은 예시이며, c의 경우 무작위에 가까운 사례로 볼 수 있다. 본 연구에서는 건축물 화재발생은 가까운 거리의 건축물 화재발생이 많았을 경우 화재발생 가능성이 높을 것이며, 가까운 건축물의 화재발생이 없었다면 화재발생이 낮았을 것이라는 것으로 가정하였다.

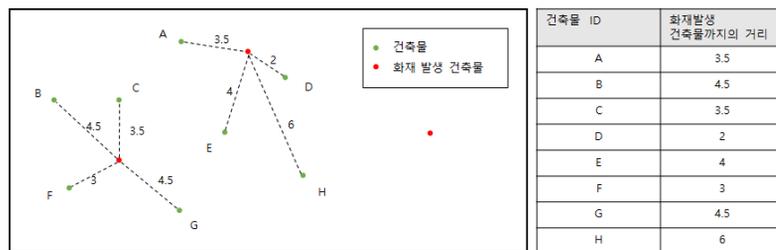


[그림 3-12] 공간자기상관의 예시

출처 : André Boraks, Gregory M. Plunkett, Thomas Morris Doro, Frazer Alo(2021). Scale-Dependent Influences of Distance and Vegetation on the Composition of Aboveground and Belowground Tropical Fungal Communities. *Microbial Ecology* 81(80):1-10.

이에 화재 발생 데이터를 두 가지 방법으로 가공하여 변수화하였다. 첫째로, 화재 발생지점과 ‘가장 근접한 화재 발생 포인트까지의 거리’를 분석하여 이를 변수화하였다. 둘째로, 건축물 주변 지역 내 화재발생 포인트 수를 집계하여 ‘화재 발생 밀도’를 나타낼 수 있도록 일정 면적당 화재발생 빈도를 분석하여 변수화하였다.

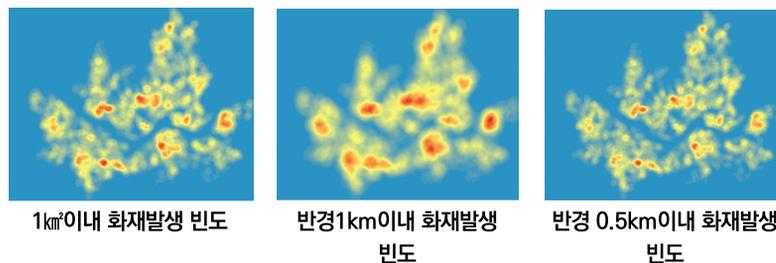
‘가장 근접한 화재 발생 포인트까지의 거리’ 변수 구축을 위해 GIS건물통합정보의 개별 건축물 위치로부터 가장 가까운 화재발생 지점까지의 거리를 분석하였다. 입력된 데이터는 건물통합정보와 화재발생 지점 데이터이며, 건물통합정보의 개별 건축물 위치를 기반으로 거리를 분석하였다. 화재발생지점 데이터는 건물통합정보_마스터 자료를 기반으로 구축되었으나, 거리분석을 위한 기준이 되는 데이터는 GIS건물통합정보를 기반으로 하여 건축물 데이터 간 거리오차가 일정부분 존재한다. 즉, 화재가 발생한 건축물은 분석 결과가 0이나 최대한 0에 가깝게 나올 것이며, 주변에 화재가 발생한 적 없는 건축물은 결과값이 크게 나타날 것이다.



[그림 3-13] 가장 근접한 화재 발생 포인트 검색 및 값 부여 개념

출처: 연구진 작성

‘화재발생 밀도’ 변수는 밀도 계산을 위한 집계대상 범위를 구분하여 총 세 가지로 구축하였다. 1km²이내 화재발생 빈도, 반경1km이내 화재발생 빈도, 반경 0.5km이내 화재발생 빈도를 집계하였다. 이에 따른 세 변수의 단위는 각각 화재 발생 포인트수 / 1km², 화재발생 포인트수 / 3.14km², 화재발생 포인트수 / 0.785km²으로 나타난다. Kernel density 기능을 사용하여 포인트 자료를 면 형태의 밀도 지도로 가공 가능하였다. 도출된 데이터는 빨간색에 가까운 픽셀일수록 밀도가 높으며, 파란색에 가까울수록 밀도가 낮도록 표출하였다.



[그림 3-14] 화재발생 밀도 분석 및 공간화 결과

출처: 연구진 작성

3) 분석 데이터 구축

□ 데이터셋 구축

화재발생 건축물 데이터셋은 화재 및 범죄발생 관련 정보와 건축물대장 관련 정보로 구성하였다. 데이터 연계를 위한 기반 데이터로는 'GIS건물통합정보' 자료를 이용하였으며, 데이터 연계는 ArcGIS 기반의 'Extract Values to Point' 분석 기법을 활용하였다. 본 연구에서 건축물 화재발생 예측을 위해 구축한 화재 및 범죄발생 관련 데이터는 모두 래스터방식의 공간정보로 구축하였다. 각 건축물이 위치한 지점의 래스터 값을 건축물에 할당하기 위해 'GIS건물통합정보'의 건축물 형상정보는 모두 포인트 데이터로 변환하여 분석을 진행하였다.

최종 구축된 화재 및 범죄 관련 변수는 10가지로, 화재 관련 변수 4개와 범죄 관련 변수 6개로 구성된다. 화재관련 변수는 자기자신을 포함한 가장 가까운 화재 발생 건축물까지의 거리, 반경 1km 및 500m 이내 화재발생 건축물 수, 주변면적 1km² 이내 화재발생 건축물 수이다. 범죄관련 변수는 m²당 강도, 강간·강제추행, 절도, 폭력, 살인, 범죄(5대 범죄 종합) 발생률의 6개이다.

[표 3-13] 화재발생 건축물 데이터셋의 화재 및 범죄관련 변수

변수명	설명	변수명	설명
NEAR_DIST	가장 가까운 화재발생 건축물까지의 거리	gg_zone	m ² 당 강간·강제추행 발생률
fire_1km	반경 1km 이내 화재발생 건축물 수	jd_zone	m ² 당 절도 발생률
fire_500m	반경 0.5km 이내 화재발생 건축물 수	pr_zone	m ² 당 폭력 발생률
fire_1km ²	1km ² 내의 화재발생 건축물 수	si_zone	m ² 당 살인 발생률
gd_zone	m ² 당 강도 발생률	crime_zone	m ² 당 범죄 발생률 (5대 범죄 종합)

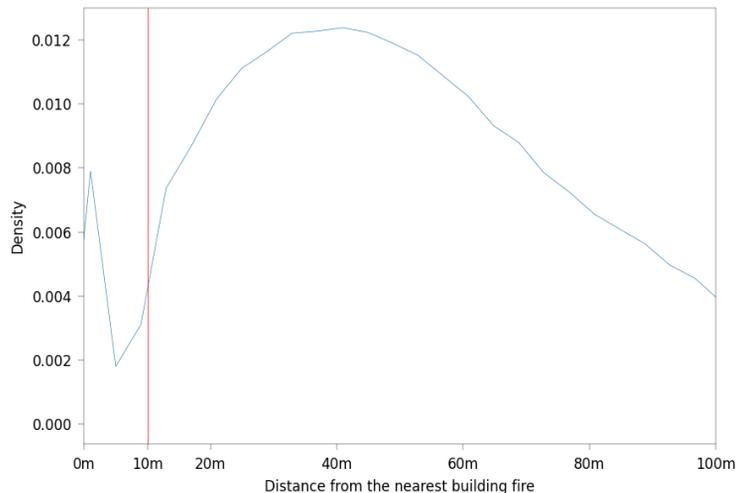
출처 : 연구진 작성

건축물대장 관련 변수는 총 6개로 건축물 용도, 구조, 연면적, 용적률, 위반건축물여부, 사용승인일을 이용하였다.

□ 화재발생 건축물 도출

최종 분석을 위한 서울시 화재발생 건축물 데이터는 '가장 근접한 화재 발생 포인트까지의 거리' 변수를 활용하여 설정하였다. 분석 결과에 따르면, 화재가 발생한 건축물은 분석 결과가 0이나 최대한 0에 가깝게 나올 것이다. 실제 화재발생 건축물을 특정하기 위해 '가장 근접한 화재 발생 포인트까지의 거리' 변수의 지시 구간별 빈도를 분석하였다.

빈도 분석에는 KDE(kernel density estimation) 분석을 사용하였다. KDE 분석은 히스토그램과 비슷하지만 연속적인 확률밀도함수를 결과로 내놓는다는 점에서 차이가 있다. 즉, 그림 3-15의 그래프는 무작위로 한 건축물을 선택하였을 때 그 건축물에서 가장 가까운 건축물 화재로부터의 거리가 얼마일지를 전체 면적이 1이 되는 확률밀도로 표현한 것이다. 예를 들어, 근접 화재로부터의 거리가 35~45m인 건축물의 비율은 구간의 폭인 10(단위 무시)과 확률밀도 약 0.012를 곱하여 전체 건축물의 약 12%에 해당함을 알 수 있다.



[그림 3-15] 가장 근접한 화재 발생 포인트까지의 KDE 분석 결과 그래프

출처 : 연구진 작성

빈도분석 결과 ‘가장 근접한 화재 발생 포인트까지의 거리’는 7m~10m 사이에서 빈도가 매우 낮은 것으로 나타났다. 이를 바탕으로 본 연구에서는 10m를 변곡점으로 설정하여 ‘가장 근접한 화재 발생 포인트까지의 거리’가 10m 이하인 건축물은 화재발생 이력이 있는 건축물로 설정하였다.

□ 학습 및 예측 데이터 구축

기계학습을 이용한 건축물 화재 예측을 위해 학습용 데이터셋과 예측용 데이터셋 구축이 필요하다. 분석을 위해 구축한 전체 데이터셋은 종속변수인 건축물의 화재발생 여부와 독립변수인 건축물대장 관련 변수 6개, 범죄관련 변수 6개, 화재관련 변수 3개로 구성된다. 즉 전체 데이터셋은 1개의 종속변수와 총 15개 독립변수로 구성하였다.

[표 3-14] 화재발생 건축물 데이터셋의 화재 및 범죄관련 변수

변수 구분		설명	
종속변수	화재여부	fire	건축물 화재발생 여부
	화재관련 변수	fire_1km	반경 1km이내 화재발생 건축물 수
		fire_500m	반경 0.5km이내 화재발생 건축물 수
fire_1km ²		1km ² 내의 화재발생 건축물 수	
독립변수	범죄관련 변수	gd_zone	㎡당 강도 발생률
		gg_zone	㎡당 강간·강제추행 발생률
		jd_zone	㎡당 절도 발생률
		pr_zone	㎡당 폭력 발생률
		si_zone	㎡당 살인 발생률
		crime_zone	㎡당 범죄 발생률 (5대 범죄 종합)
독립변수	건축물 관련 변수	용도	건축물 용도 코드
		구조	건축물 구조 코드
		연면적	건축물 연면적
		용적률	건축물 용적률
		위반여부	위반건축물여부
		사용승인	건축물 사용승인 연도

출처 : 연구진 작성

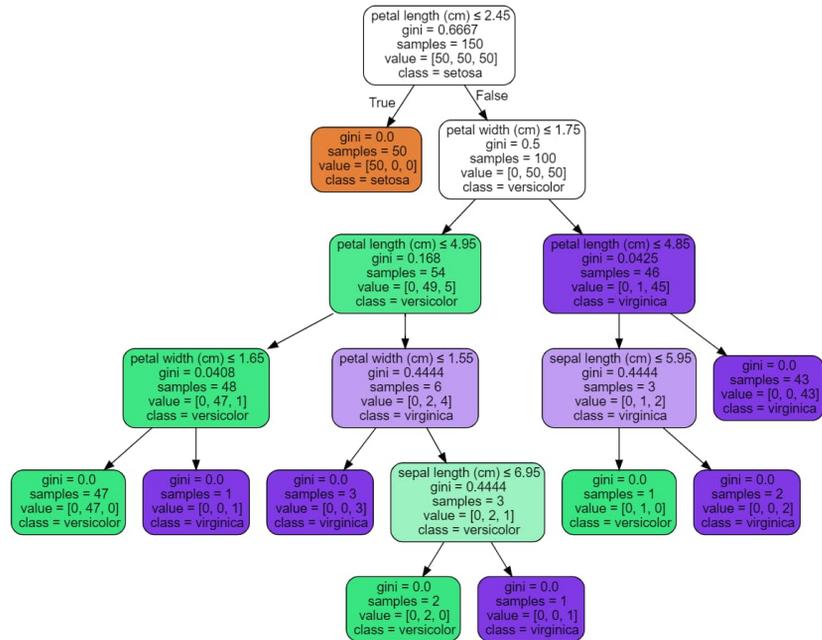
4) 기계학습 모델 선정 및 분석 데이터 전처리

① 기계학습 모델 선정

재난·재해 취약 건축물 예측을 위한 화재 데이터 학습 모델로는 의사결정나무와 딥러닝 모델인 TabNet 두 가지 모델을 선정하여, 해석가능성과 예측 성능 측면에서 장단점이 서로 보완될 수 있도록 하였다. 의사결정나무는 모델의 결과 예측을 해석할 수 있는 몇 안 되는 기계학습 모델로, 모델의 예측력은 떨어지더라도, 예측 결과가 도출된 원인을 사람이 쉽게 파악할 수 있다는 장점이 있다. 반면 딥러닝은 기계학습 모델 중에서도 특히 복잡도가 높아 원인 해석이 어려우나, 다양한 분야에서 최고 수준의 성능을 보이고 있고 새로운 모델 연구도 활발하게 진행되고 있어 분석 방법론에 적용하였다.

다양한 의사결정나무 모델 중에서, 본 연구에서는 scikit-learn에서 구현한 CART (Classification and Regression Trees) 모델을 사용하였다. CART는 전통적으로 널리 사용되는 C4.5 모델과 유사하나, 수치형(numerical) 종속 변수를 취급할 수 있어 분류(classification) 뿐만 아니라 회귀(regression) 문제에도

적용할 수 있다는 장점이 있다. 본 연구에서는 화재 발생 여부(fire 변수)가 이진 (binary) 변수인 점을 고려하여 분류 모델을 사용하였다. 단, scikit-learn의 CART 구현은 범주형 변수를 다루지 못하기 때문에 그에 대한 전처리가 필요하였다.²¹⁾



[그림 3-16] scikit-learn의 의사결정나무(CART) 모델 분석 결과 예시

출처 : scikit-learn developers(2022). Decision Trees. Scikit-Learn. <https://scikit-learn.org/stable/modules/tree.html> (검색일 : 2022.10.18.)

화재 및 범죄 데이터 학습을 위하여 선정된 딥러닝 모델은 테이블 형태의 정형 데이터 분석을 수행하는 모델인 TabNet이다. TabNet은 정형 데이터 분석을 위하여 의사결정나무 및 앙상블 방법론의 여러 아이디어를 반영한 딥러닝 모델로, 의사결정나무 기반 앙상블 모형에 못지않은 성능을 보이면서도, 데이터에서 학습한 규칙에 대한 해석가능성도 고려한 모델이다.²²⁾ 딥러닝의 발전 가능성을 고려하여, 본 연구에서는 TabNet을 통해 딥러닝 기반 건축물 빅데이터 분석의 가능성을 검토하고자 하였다.

21) scikit-learn developers(2022). Decision Trees. Scikit-Learn. <https://scikit-learn.org/stable/modules/tree.html>

22) Arik, S. O., & Pfister, T. (2020). TabNet: Attentive Interpretable Tabular Learning (arXiv:1908.07442). arXiv. <https://doi.org/10.48550/arXiv.1908.07442>

② 분석 데이터 전처리

기계학습 모형과 구현체에 따라 처리할 수 있는 입력 데이터의 형식이 조금씩 다르기 때문에, 분석 데이터를 전처리(preprocessing)하여 입력 가능한 형식으로 변환할 필요가 있다. 따라서 기계학습 모델 선정 후 분석 대상 데이터를 각 모델에 맞추어 전처리하는 과정을 거쳐 기계학습 모델에 입력할 데이터를 준비하였다.

□ 분류 범주 불균형 해결을 위한 언더샘플링

기계학습에 종속변수인 분류 범주가 불균형한 데이터를 사용할 경우 학습이 정상적으로 이루어지지 않으며, 이를 클래스 불균형 문제(class imbalance problem)라고 한다. 건축물 화재는 드물게 일어나는 사건으로, 분석 대상 데이터에서 최근 5년간 화재가 발생한 건축물은 540,111동 중 20,480동, 3.8%에 불과하였다.

본 연구에서는 분류 범주가 균형 잡힌 데이터를 마련하기 위하여 언더샘플링(undersampling) 방법을 사용하였다. 구체적으로, 화재가 발생한 건축물 전수와 화재가 발생하지 않은 건축물 중 같은 수만큼 무작위 추출한 표본을 1:1로 섞어, 총 40,960동의 데이터를 분석 대상 데이터로 마련하였다.

□ 데이터셋 분할

기계학습 모델은 데이터에 잠재된 규칙을 찾아낼 수 있지만, 학습을 계속 진행하면 주어진 데이터에만 적용되는 특수한 규칙을 찾아내어 실제 현상에 대한 예측 정확도가 낮아지는 과적합(overfitting) 현상이 발생한다. 기계학습에서는 이러한 현상을 방지하기 위하여 기계학습 모델이 본 적 없는 데이터에 대한 정확도를 주기적으로 측정하는데, 이를 위해서는 기계학습 모델에 노출시키지 않을 데이터를 별도의 데이터셋으로 분리할 필요가 있다.

본 연구는 분석 대상 데이터를 무작위 표본 추출로 분할하여, 80%는 훈련(train) 데이터로 예측모형 학습에 활용하고, 10%는 검증(validation) 데이터로 모형 복잡도 결정 및 훈련 조기 중단(early stopping)에 활용하였으며, 10%는 시험(test) 데이터로 예측모형의 최종 정확도 평가 및 학습된 규칙 해석에 활용하였다.

□ 범주형 변수 이진화

scikit-learn의 의사결정나무 모델 구현은 범주형 변수를 처리할 수 있는 CART 알고리즘을 사용하지만, 현재는 범주형 변수가 지원되지 않고 있다. 따라서 범주형 변수를 연속형 변수처럼 취급할 수 있도록 이진형 변수로 변환하였다.

범주형 변수의 변환에는 원핫 인코더(one-hot encoder)를 사용하였다. 원핫 인코더는 여러 가지 값을 가질 수 있는 범주형 변수를 0 또는 1의 값을 갖는 이진형 변수 여러 개로 변환한다. 각 레코드는 해당하는 범주에 따라 한 변수만 1의 값을 갖고, 해당하지 않는 범주에 대한 변수는 모두 0의 값을 갖게 되며, 원핫이라는 이름도 이와 같은 처리 후 데이터의 모습을 지칭한다.

[표 3-15] 원핫 인코더 적용 예시

용도	용도코드		용도_01000	용도_02000	용도_03000	...	용도_27000	용도_None
단독주택	01000	→	1	0	0	...	0	0
공동주택	02000	→	0	1	0	...	0	0
...
관광휴게 시설	27000	→	0	0	0	...	1	0
공동주택	02000	→	0	1	0	...	0	0

출처 : 연구진 작성

□ 결측값 대체

건축물과 관련한 다양한 출처의 데이터가 종합된 분석 대상 데이터는 일부 데이터가 존재하지 않고 대신 데이터의 부재를 나타내는 값으로 채워져 있는데, 이를 결측값(null)이라고 한다. DreamQuark의 TabNet은 이러한 결측값을 처리하지 못하기 때문에 다른 값으로 대체해야 한다.

범주형 변수의 경우, 결측값을 별도 범주인 것처럼 처리하면 쉽게 처리가 가능하다. 원핫 인코더에서도 결측값에 대한 특별한 배려 없이 해당 값을 여러 범주 중 하나로 처리하는데, [표 3-15]의 '용도_None'이 이와 같은 처리 결과로 나타난 것이다.

연속형 변수의 결측값은 존재하는 데이터의 평균값을 적용하여 처리할 수 있다. 다만, 전체 데이터셋의 평균을 적용하면 데이터셋 사이에 정보가 교류하게 되므로, TabNet 모델이 훈련 과정에서 검증 데이터와 시험 데이터의 영향을 받게 된다. 따라서 각 데이터셋별로 평균을 내어 각각의 평균값을 결측값에 적용하였다.

5) 기계학습을 활용한 화재 취약 건축물 예측

① 기계학습을 활용한 화재 취약 건축물 예측

앞에서 선정한 의사결정나무 및 TabNet 모델과 이에 맞게 전처리한 데이터를 활용하여, 개별 건축물의 최근 5년 화재 발생 여부를 예측하도록 하고, 그 결과를 통하여 건축물의 화재 취약성을 평가할 수 있는 방법론을 구축하였다.

화재가 발생한 건축물은 극히 드문 수준이므로 화재 발생 여부는 매우 불균형하여 예측하기 어려운 변수이다. 그러나 이러한 예측을 수행하는 목적은 이미 화재가 발생한 건축물을 예측하는 데 사용되는 여러 측면의 특징에서 공통점을 지니는 건축물은 마찬가지로 화재에 취약할 것이라는 가정 하에, 화재 발생과 연관된 공통점이 있는 화재 취약 건축물을 찾기 위한 것이다. 따라서 이 경우에는 최근 5년 동안 화재가 발생하지 않았으나 기계학습 모델이 화재가 발생한 건축물이라고 (잘못) 판단한 건축물 목록이 화재 취약 건축물 예측의 실제 결과가 된다.

② 의사결정나무 모델을 활용한 화재 취약 건축물 유형 추출

의사결정나무 모델은 데이터에서 학습된 규칙이 사람이 이해하기 쉬운 순서도 형태로 표출되어 쉽게 해석이 가능하다. 본 연구에서는 건축물 화재 발생 여부 예측을 학습한 의사결정나무의 구조를 통하여 화재 취약 건축물의 유형을 추출하고, 건축물 속성 및 범죄 등 다양한 변수가 화재 위험성과 보이는 연관성을 분석하고자 하였다.

③ TabNet 모델을 활용한 화재 위험 영향 요인 추출

딥러닝 모델인 TabNet은 딥러닝 모델 특유의 문제인 해석의 어려움을 해결하고자 모델에 대한 자체적인 해석 방법론을 직접 제시하였다. TabNet 모델은 학습한 예측 기준에 따른 변수의 중요도, 각 판단 단계에서 나타나는 변수 간의 연관성 등에 대한 해석이 가능하다.

본 연구에서는 건축물 화재 발생 여부 예측을 학습한 TabNet 모델에서 건축물 화재 위험 영향 요인 등을 추출해보고, 단순 작업 수행이 아닌 학습 결과 해석이 필요한 건축물 데이터 분석에서 딥러닝 모델의 적용 가능성을 검토하고자 한다.

3. 소결

본 연구는 사회재난 중 건축물에서 빈번하게 발생하고 사고위험이 높으며 국민들이 불안감이 높은 화재 발생 위험이 높은 건축물을 예측하는 모델을 개발하는 것을 주요 목적으로 하고 있다.

건축물 화재 예측 모델 구축을 위해서는 우선적으로 대상지 선정 시 다양한 요소를 고려하여야 한다. 공간적 범위가 넓어야 하고, 화재발생 데이터의 경우 빅데이터라 할 수 있고 기계학습을 수행할 수 있을 정도로 발생 건수가 많아야 하며, 도시 내 다양한 인문사회학적 요소가 집합되어 범용적 특징을 보이는 곳을 선정하여야 한다. 이에 본 연구에서는 서울시를 그 대상으로 선정하였다.

연계 빅데이터 구축을 위하여 분석 대상인 화재발생 데이터와 건축행정 데이터를 연계하였다. 먼저 화재발생 데이터와 건축행정 데이터의 구조를 확인하고, 두 데이터의 연계 KEY로 주소정보를 담고 있는 PNU 코드를 생성하였다. 이를 통해 화재가 발생한 건축물의 속성정보를 취득할 수 있다. PNU코드로 화재데이터와 건축행정데이터를 연계시키는 경우 건축물은 그 속성에 따라 단일건축물과 2개동 이상의 집합건축물로 구분되기 때문에, PNU 코드로 연계한 후 상세주소 확인을 통해 화재가 발생한 건축물의 동을 확인해야 한다.

화재데이터와 건축행정데이터 연계 후에는 빅데이터 분석을 위한 거리변수 도입을 위하여 공간정보화 즉 지오코딩을 진행한다. 이는 주요변수로 활용될 다른 데이터도 동일하다. 지오코딩은 주소정보를 기반으로 진행하며, 주의할 점은 지오코딩 시 지적이 기준이기 때문에 건축행정데이터와 연계할 때와 마찬가지로 정확한 동을 특정하여야 한다. 이에 본 연구에서는 건물통합정보를 Basemap으로 활용하여 진행하였다. 본 연구에서 변수로 도입할 범죄데이터의 경우 자료가 방대하기 때문에 범죄발생 건축물을 추출하는 대신 집계구 단위로 도출하였다.

공간정보화를 진행한 이후에는 기계학습을 위해 화재 데이터를 가공하고 공간변수를 구축한다. 본 연구에서는 건축물의 화재발생은 공간 자기상관이 있을 것으로 가정하기 때문에 거리변수를 구축하였다. 변수 구축 이후에는 최종 데이터셋을 구축하였다. 데이터셋 구축을 위해 ArcGis프로그램을 활용하였으며, 화재관련 변수 3종, 범죄관련 변수 6종, 건축물 관련 변수 6종을 독립변수로 구축하고 건축물 화재발생 여부를 종속변수로 구축하였다.

빅데이터 분석을 위한 변수를 설정한 후 해당 변수를 활용하여 화재 위험지역 예측모델 구축을 위한 기계학습 모델을 선정하였다. 본 연구에서는 해석가능성과

예측 성능 측면에서 장단점이 보완될 수 있도록 의사결정나무와 딥러닝 모델을 병행하여 사용하는 것을 제안하였다. 구체적인 모델은 의사결정나무 모델 중에서 화재발생 여부가 이진변수인 점을 고려하여 scikit-learn에서 구현한 CART 모델, 딥러닝 모델 중에서 데이터가 테이블 형태임을 감안하여 TabNet모델 중 pyTorch 기반 pytorch-tabnet을 제안하였다.

기계학습 모델을 선정한 이후에는 분석 대상 데이터를 각 모델에 맞추어 전처리 하는 과정을 거쳐야 한다. 본 연구에서 활용하는 최근 5년간의 화재발생 건축물 동수는 20,480동으로 서울시 전체 건축물 동수의 3.8%에 불과하기 때문에 언더 샘플링 방법을 사용하여 화재발생 건축물 수와 동일한 수의 건축물을 무작위 추출하여 총 40,960동의 데이터를 최종 분석 대상 데이터로 추출하였다. 다음으로 분석 대상 데이터를 무작위 표본 추출로 분할하여 80%는 훈련데이터로 활용하고, 10%는 검증데이터, 10%는 시험데이터로 활용하고자 한다.

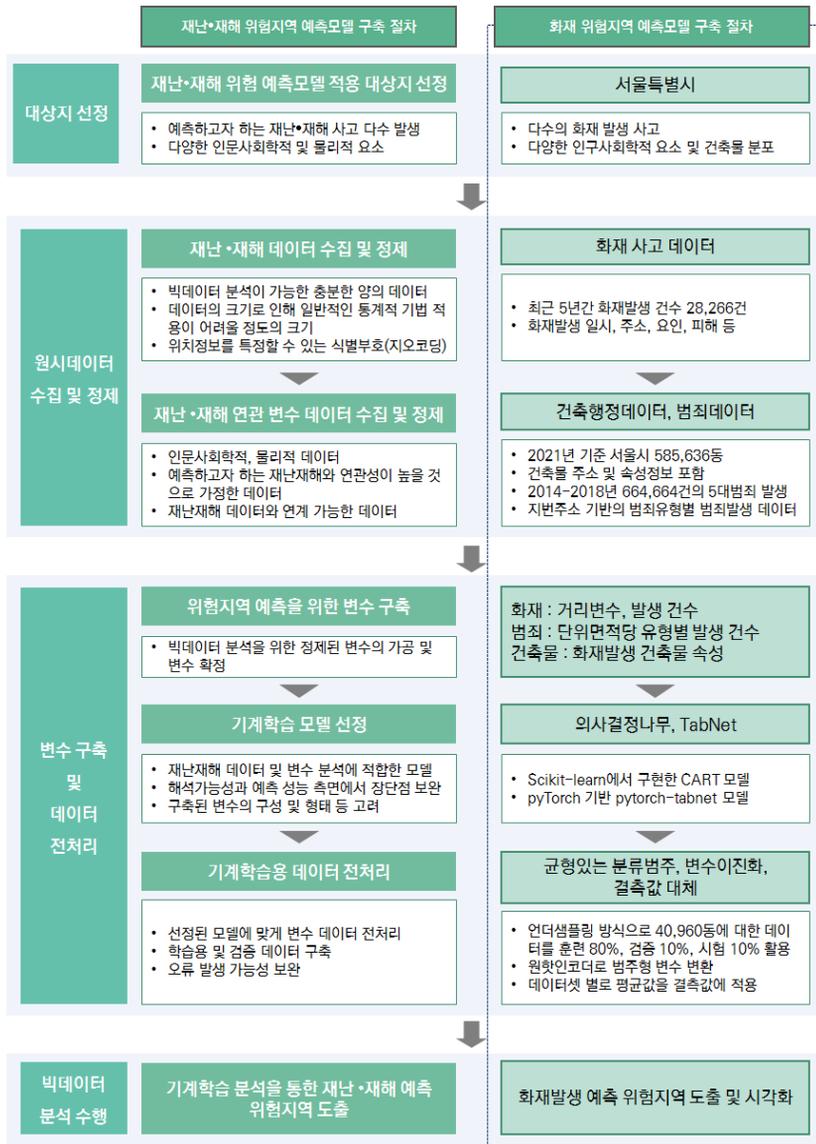
데이터셋을 분할한 후에는 범주형 변수 이진화를 수행한다. 본 연구에서는 원핫 인코더로 범주형 변수를 변환하였다. 이를 통해 각 레코드는 해당하는 범주에 따라 한 변수만 1의 값을 갖고, 해당하지 않는 범주에 대한 변수는 모두 0의 값을 갖게 된다. 마지막으로 일부 데이터가 존재하지 않아 데이터의 부재를 나타내는 값으로 채워지는 결측값을 대체하여야 하며, 본 연구에서는 각 데이터셋 별로 평균을 내어 각각의 평균값을 결측값에 적용하였다. 이와 같은 과정을 거쳐 데이터 전처리를 완료하였다.

기계학습을 위한 데이터 가공, 변수구축, 전처리가 완료되면 앞서 선정한 분석 모델을 통해 위험지역 예측을 수행한다. 본 연구에서는 의사결정나무 모델을 적용하여 화재 취약 건축물의 유형을 추출하고, 건축물 속성 및 범죄 등 다양한 변수와 화재 위험성과의 연관성을 분석한다. TabNet모델을 통해서서는 건축물 화재 위험 영향 요인 등을 추출하고, 결과적으로 기계학습 결과의 해석을 기반으로 건축물 데이터 분석에 있어 딥러닝 모델의 적용 가능성을 검토한다.

상기한 내용은 일련의 빅데이터 구축 및 가공, 분석에 대한 시나리오로 이를 정리 하면 빅데이터 기반 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델로 정리할 수 있다.

재난·재해 위험지역 예측 방법론은 1. 대상지 선정- 2. 분석을 위한 데이터 수집- 3. 변수 구축을 위한 데이터 가공- 4. 변수구축- 5. 빅데이터 분석 모델 선정- 6. 모델에 적합하도록 변수 전처리- 7. 분석을 통한 예측으로 정리할 수 있다. 건축물 화재 예측 모델은 재난·재해 위험지역 예측 방법론의 절차에 맞추어 관련

내용을 화재 예측에 맞추어 종합하는 방식으로 개발하였다. 세부적인 구축 절차는 다음과 같다.

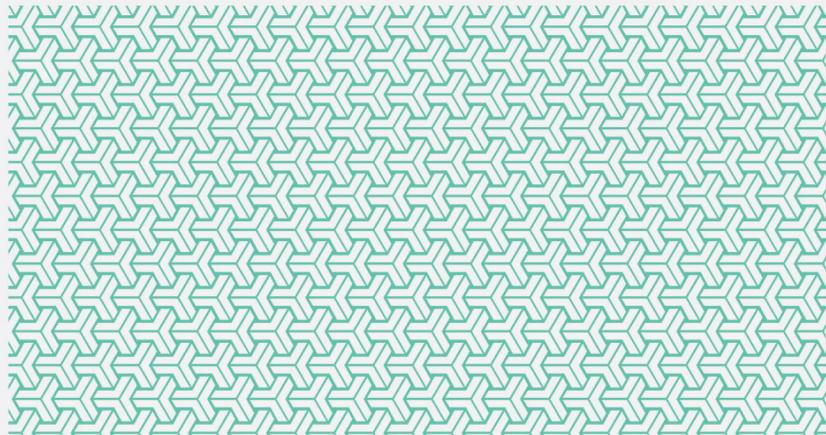


[그림 3-19] 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델

출처 : 연구진 작성

제4장 건축물 화재 예측 모델 적용

Chapter.4



1. 건축물 특성에 따른 화재 통계
2. 기계학습 기반 화재 취약 건축물 예측
3. 소결 및 시사점

1. 건축물 특성에 따른 화재 통계

기계학습 기반 건축물 화재 예측 분석에 앞서 예측모델 개발 과정에서 구축된 서울시 건축물 화재 연계 데이터를 활용한 기술통계를 생산하였다. 화재 발생 빈도 통계는 건축물 용도별, 연면적별, 층수별, 노후도별로 분석하였으며, 화재피해는 구/동별, 구별인구 및 면적대비, 지상 층수 및 지하 층수, 건축물 용도별, 건축 승인년도 별로 분석하였다. 또한 화재피해에 대해서는 이중결합 통계((화재발생 정보 + 건축물대장정보)로서 건축물 용도와 승인년도, 건축물 지상층수와 승인년도 화재 피해를 분석하였다.

① 서울시 건축물 용도·연면적·층수·노후도 별 화재 발생 통계

본 연구에서의 서울시 건축물 화재 발생 통계는 전체 화재발생 건축물만을 대상으로 통계를 기술하였다. 우선 건축물 용도별 화재발생 통계를 살펴보면 최근 5년(2017년~2021년) 서울시 화재 전체 21,367건 중에서 공동주택에서의 화재 발생건수가 5,114건(23.93%)으로 가장 많았으며, 다음으로 단독주택(단독, 다가구)이 4,902건(22.94%)으로 높았다. 제1종 및 제2종 근린생활시설에서는 총 6,180건(28.92%)의 화재가 발생하였다.

[표 4-1] 최근 5년간(2017~2021년) 서울시 건축물 용도별 화재 발생 건수

구분	2017	2018	2019	2020	2021	합계	비율
공동주택	1,109	1,271	1,103	939	692	5,114	23.93%
단독주택	1,128	1,222	1,066	879	607	4,902	22.94%
제2종 근린생활시설	745	768	699	590	483	3,285	15.37%
제1종 근린생활시설	614	703	647	566	365	2,895	13.55%
업무시설	242	252	251	260	182	1,187	5.56%
교육연구시설	56	72	73	51	57	309	1.45%
공장	62	49	55	42	25	233	1.09%
숙박시설	35	55	50	38	28	206	0.96%
판매시설	29	44	52	32	22	179	0.84%
노유자시설	34	32	27	19	18	130	0.61%
기타	73	80	87	82	53	375	1.76%
미기재	695	600	528	413	316	2,552	11.94%
합계	4,822	5,148	4,638	3,911	2,848	21,367	100%

출처 : 서울시 내부자료를 활용하여 연구진 작성

건축물 연면적별 화재 발생 통계를 살펴보면 연면적 200㎡초과 500㎡이하의 건축물에서 총 5,192건(24.3%)로 가장 많이 발생하였으며, 연면적 500㎡초과 1,000㎡이하에서는 4,108건(19.23%)의 화재가 발생하였다. 건축물 연면적 화재 발생 통계를 살펴보면 일반적으로 연면적 1,000㎡이하 또는 3,000㎡이상의 건축물에서 화재 발생 빈도가 높음을 알 수 있다.

[표 4-2] 최근 5년간(2017~2021년) 서울시 건축물 연면적별 화재 발생 건수

구분	2017	2018	2019	2020	2021	합계	비율
200㎡ 이하	705	809	669	549	399	3,131	14.65%
200㎡초과 500㎡이하	1,174	1,224	1,139	996	659	5,192	24.30%
500㎡초과~ 1,000㎡이하	915	976	904	735	578	4,108	19.23%
1,000㎡초과~ 2,000㎡이하	433	465	406	371	269	1,944	9.10%
2,000㎡초과~ 3,000㎡이하	158	180	181	146	107	772	3.61%
3,000㎡초과	742	894	811	701	520	3,668	17.17%
미기재	695	600	528	413	316	2,552	11.94%
합계	4,822	5,148	4,638	3,911	2,848	21,367	100.0%

출처 : 서울시 내부자료를 활용하여 연구진 작성

건축물 층수별 화재 발생 통계를 살펴보면 5층 이하 건축물에서의 화재 발생 건수가 13,964건(65.35%)로 매우 높았다. 이는 5층 이하 건축물이 서울시 전체 건축물에서 차지하는 비중이 높기 때문인 것으로 보인다. 다음으로 6~10층 건축물의 경우 2,257건(10.56%)로 나타났다.

[표 4-3] 최근 5년간(2017~2021년) 서울시 건축물 층수별 화재 발생 건수

구분	2017	2018	2019	2020	2021	합계	비율
5층 이하	3,160	3,390	3,051	2,555	1,808	13,964	65.35%
6~10층	461	518	491	445	342	2,257	10.56%
11~20층	403	496	466	399	308	2,072	9.70%
21~30층	81	118	86	79	62	426	1.99%
30층이상	22	26	16	20	12	96	0.45%
미기재	695	600	528	413	316	2,552	11.94%
합계	4,822	5,148	4,638	3,911	2,848	21,367	100%

출처 : 서울시 내부자료를 활용하여 연구진 작성

서울시 노후도별 화재발생 통계를 살펴보면 30년 이상 건축물의 화재 발생 건수가 총 8,102건(37.92%)으로 건축물이 노후화 될수록 화재 발생 건수가 높다는 것을 알 수 있다. 다음으로 21~30년 사이의 건축물에서는 4,806건(22.49%)의 화재가 발생하였다.

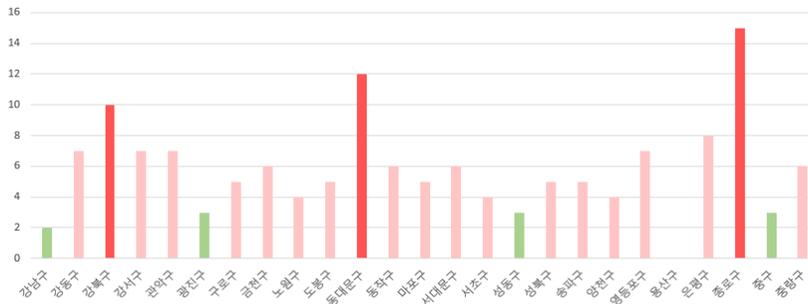
[표 4-4] 최근 5년간(2017~2021년) 서울시 건축물 노후도별 화재 발생 건수

구분	2017	2018	2019	2020	2021	합계	비율
10년 이하	295	429	481	453	340	1,998	9.35%
11년~20년	765	846	741	616	511	3,479	16.28%
21년~30년	1,113	1,125	1,111	850	607	4,806	22.49%
31년~40년	1,159	1,276	1,037	940	631	5,043	23.60%
41년 이상	701	776	642	559	381	3,059	14.32%
기타	789	696	626	493	378	2,982	13.96%
합계	4,822	5,148	4,638	3,911	2,848	21,367	100%

출처 : 서울시 내부자료를 활용하여 연구진 작성

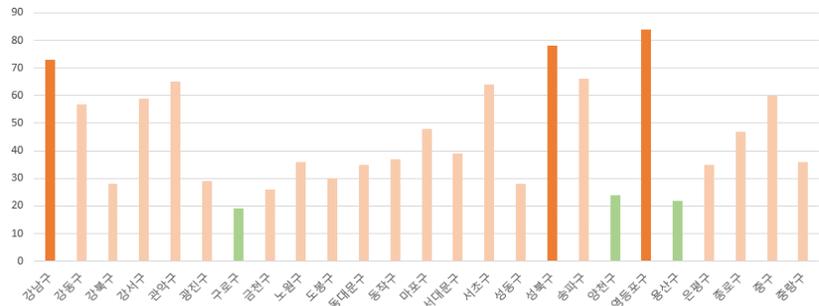
② 서울시 구/동별 화재 통계

서울시 구별, 동별 화재 사망자 수, 부상자 수, 인명피해, 재산피해액 합계 등은 다음과 같다. 서울시 구별 화재 사망자 수 상위 세 곳은 종로구, 동대문구, 강북구 순이며, 하위 지역은 용산구, 강남구, 광진구/성동구/중구 순으로 나타난다. 서울시 구별 화재 부상자 수 상위 세 지역은 영등포구, 성북구, 강남구이며, 하위 세 지역은 구로구, 용산구, 양천구 순으로 나타났다.



[그림 4-1] 서울시 구별 화재 사망자 수 합계

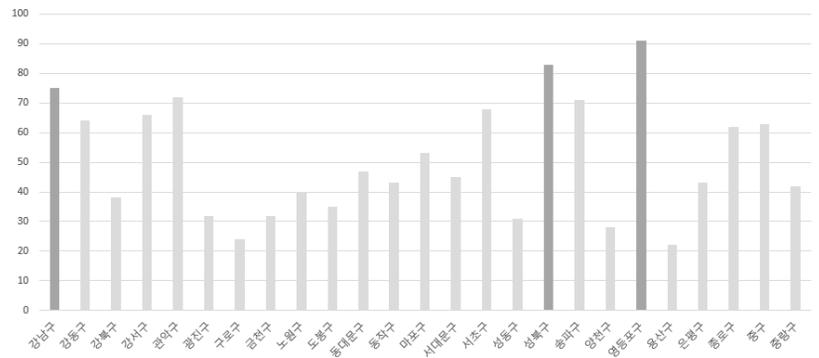
출처 : 서울시 내부자료를 활용하여 연구진 작성



[그림 4-2] 서울시 구별 화재 부상자 수 집계

출처 : 서울시 내부자료를 활용하여 연구진 작성

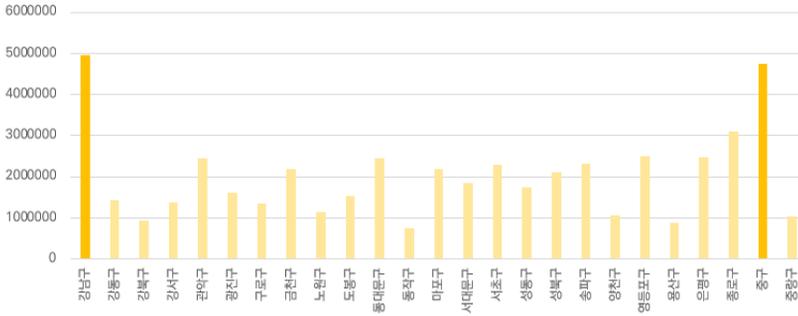
서울시 구별 인명피해 상위 세 지역은 영등포구, 성북구, 강남구이며, 하위 세 지역은 용산구, 구로구, 양천구 순으로 나타났다. 인명피해는 사망자 수와 부상자 수의 합계를 나타낸다.



[그림 4-3] 서울시 구별 인명피해 집계

출처 : 서울시 내부자료를 활용하여 연구진 작성

서울시 구별 재산피해 소계의 경우, 중구에서 이상치가 탐지되어 기초통계 분석을 진행하기 전 데이터 전처리의 과정으로 outlier를 제거하여 분석을 진행하였다. 그 결과, 서울시 구별 재산피해 소계를 살펴보면, 강남구, 중구 순으로 가장 피해액이 높았으며, 동작구, 용산구 순으로 피해액이 낮은 것으로 보인다. 강남구 중에서는 역삼동이, 중구 중에서는 신당동이 높은 재산피해를 보이고 있으며, 동작구 중에서는 동작동, 용산구 중에서는 용산동 5가 지역이 낮은 재산피해액을 나타내고 있음을 파악하였다.



[그림 4-4] 서울시 구별 재산피해 합계

출처 : 서울시 내부자료를 활용하여 연구진 작성

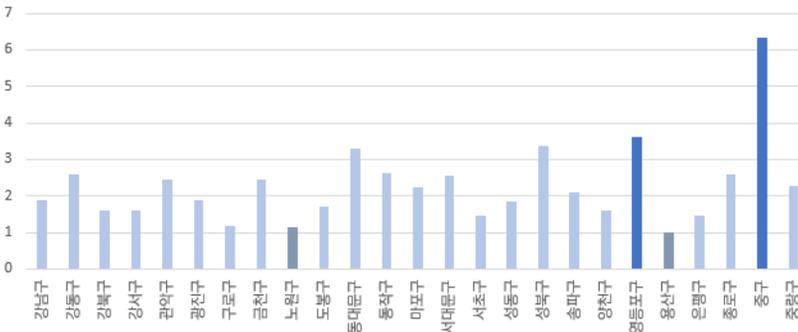
③ 서울시 구별 인구와 면적 대비 화재 인명피해

서울시 구별 인구수와 면적을 반영하여 화재로 인한 인명피해를 정리하였다. 인구수 대비 인명피해가 높은 두 지역은 종로구와 중구이며, 적은 두 지역은 양천구와 구로구로 나타난다. 면적 대비 인명피해가 높은 두 지역은 중구와 영등포구이며, 적은 두 지역은 용산구와 노원구로 나타난다.



[그림 4-5] 서울시 구별 인구수 대비 인명피해

출처 : 서울시 내부자료를 활용하여 연구진 작성



[그림 4-6] 서울시 구별 면적 대비 인명피해

출처 : 서울시 내부자료를 활용하여 연구진 작성

서울시 구와 동별 사망자 수 / 부상자 수 / 인명피해 / 재산피해를 함께 정리한 표는 아래와 같다.

[표 4-5] 서울시 구별 인구수/면적 대비 인명피해

구	인명피해	인구수	면적(km ²)	인구수대비 인명피해	면적 대비 인명피해
종로구	62	151,999	23.91	0.000408	2.593057
중구	63	130,764	9.96	0.000482	6.325301
용산구	22	233,734	21.87	0.000094	1.005944
성동구	31	288,941	16.86	0.000107	1.838671
광진구	32	351,073	17.06	0.000091	1.875733
동대문구	47	353,181	14.22	0.000133	3.305204
중랑구	42	391,110	18.50	0.000107	2.270270
성북구	83	442,596	24.58	0.000188	3.376729
강북구	38	298,622	23.60	0.000127	1.610169
도봉구	35	314,884	20.65	0.000111	1.694915
노원구	40	509,686	35.44	0.000078	1.128668
은평구	43	474,124	29.71	0.000091	1.447324
서대문구	45	317,159	17.63	0.000142	2.552467
마포구	53	376,062	23.85	0.000141	2.222222
양천구	28	445,270	17.41	0.000063	1.608271
강서구	66	575,813	41.45	0.000115	1.592280
구로구	24	416,570	20.12	0.000058	1.192843
금천구	32	243,872	13.02	0.000131	2.457757
영등포구	89	398,938	24.55	0.000223	3.625255
동작구	43	392,628	16.36	0.000110	2.628362
관악구	72	500,750	29.57	0.000144	2.434900
서초구	68	409,025	46.98	0.000166	1.447424
강남구	75	534,988	39.50	0.000140	1.898734
송파구	71	665,847	33.88	0.000107	2.095632
강동구	64	464,490	24.59	0.000138	2.602684

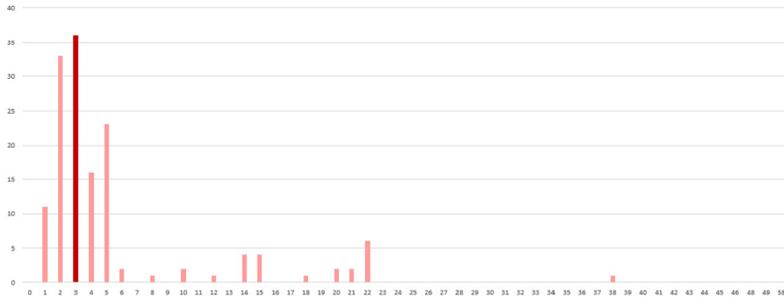
출처 : 인구수 - <https://data.seoul.go.kr/dataList/419/S/2/datasetView.do>(검색일 : 2022.09.10.)

면적 - <https://data.seoul.go.kr/dataList/412/S/2/datasetView.do>(검색일 : 2022.09.10.)

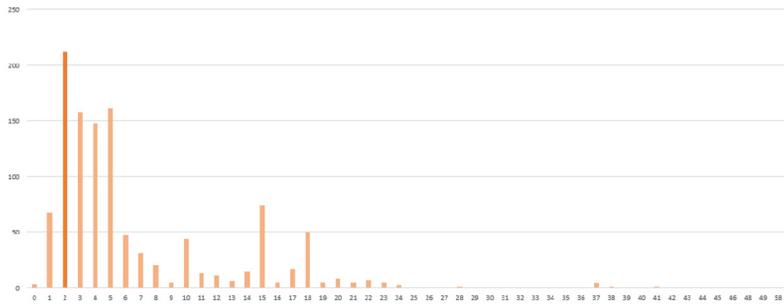
④ 건축물 층수별 화재 피해

화재 발생과 건축물의 특성을 함께 비교하기 위해 지상 층수 및 지하 층수에 따른 화재 피해 정도를 파악하였다. 건물의 층수에 따라 화재 발생 시 피해 정도가 다를 것으로 예상되어 분석을 진행하였다.

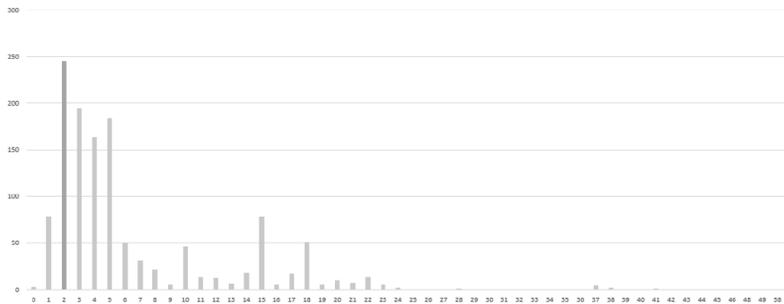
지상 층수에 따른 화재로 인한 사망자 수, 부상자 수, 인명피해, 재산피해 합계는 아래와 같으며, 주로 저층 건축물의 피해정도가 높은 것으로 나타나는데, 이는 총 합계로서 계산되었기 때문으로 추정된다.



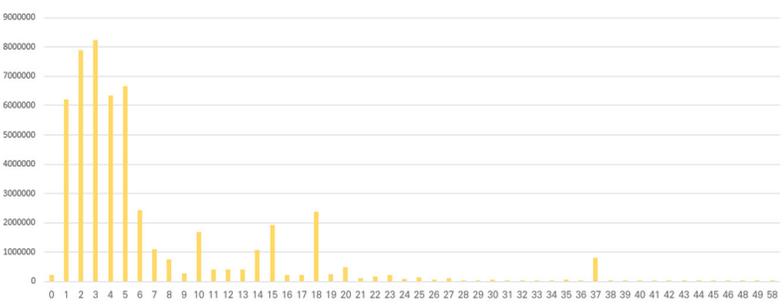
[그림 4-7] 지상 층수에 따른 합계 사망자 수
출처 : 서울시 내부자료를 활용하여 연구진 작성



[그림 4-8] 지상 층수에 따른 합계 부상자 수
출처 : 서울시 내부자료를 활용하여 연구진 작성

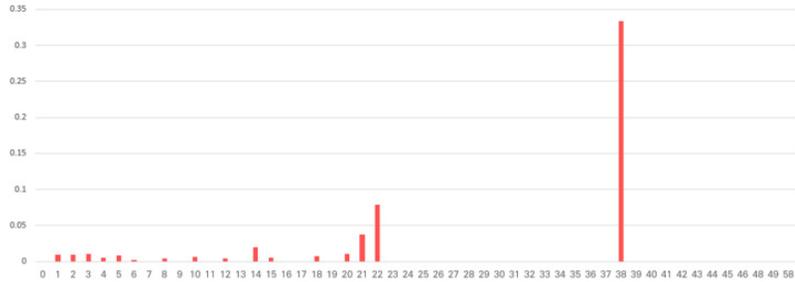


[그림 4-9] 지상 층수에 따른 합계 화재 인명피해
출처 : 서울시 내부자료를 활용하여 연구진 작성



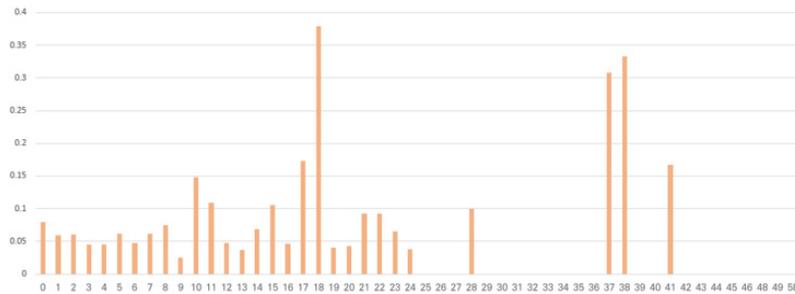
[그림 4-10] 지상 층수에 따른 합계 화재 재산피해
출처 : 서울시 내부자료를 활용하여 연구진 작성

총 합계로 계산하면 발생 빈도수에 따라 영향력이 달라지기 때문에 화재가 발생한 건축물의 동수 평균으로 계산하는 것이 더 바람직 할 것으로 예상된다. 따라서 지상 층수에 따른 화재 사망자 수, 부상자 수, 인명피해, 재산피해의 평균값을 계산하여 재분석하였다. 합계로 계산 되어진 것과는 조금 다른 결과가 나타났다.



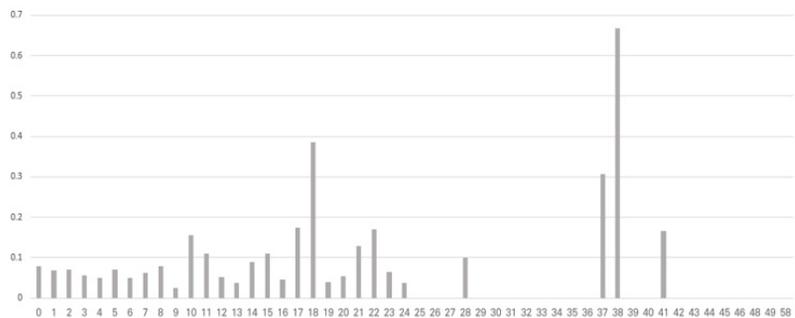
[그림 4-11] 지상 층수에 따른 평균 사망자 수 히스토그램

출처 : 서울시 내부자료를 활용하여 연구진 작성



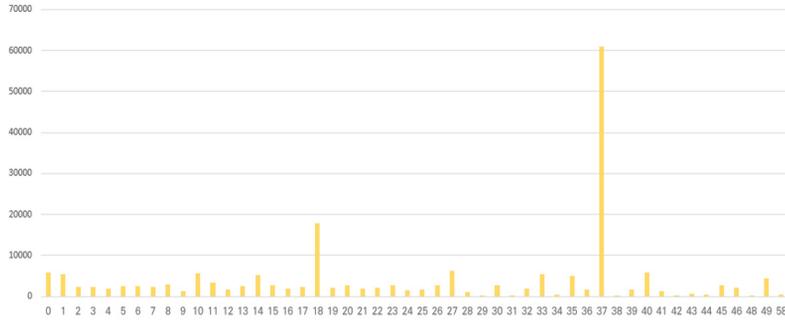
[그림 4-12] 지상 층수에 따른 평균 부상자 수 히스토그램

출처 : 서울시 내부자료를 활용하여 연구진 작성



[그림 4-13] 지상 층수에 따른 평균 인명피해 히스토그램

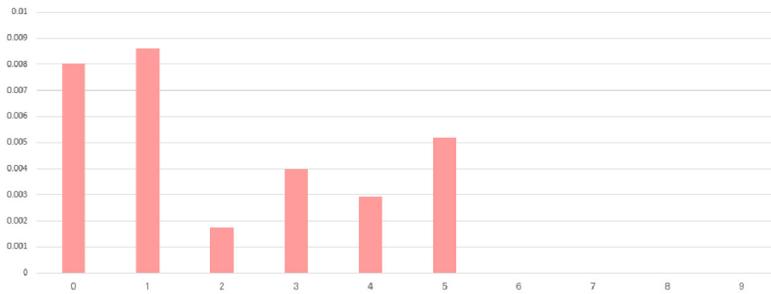
출처 : 서울시 내부자료를 활용하여 연구진 작성



[그림 4-14] 지상 층수에 따른 평균 재산피해 히스토그램

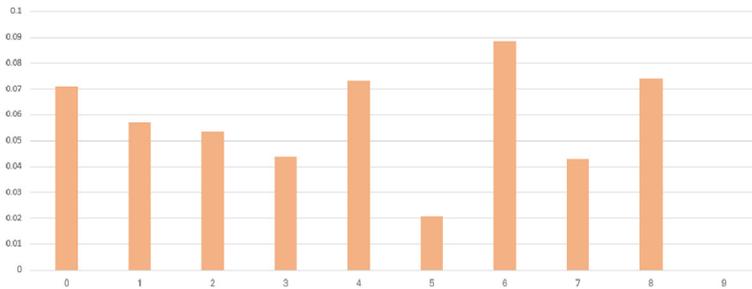
출처 : 서울시 내부자료를 활용하여 연구진 작성

결과를 살펴보면 평균 사망자 수, 부상자 수, 인명피해, 재산피해 4가지 항목에 대해서 대체로 지상 층수가 높은 건물일수록 더 많은 피해를 입었음을 알 수 있다. 고층 건물일수록 화재 발생 시 진압에 소요되는 시간도 길고, 제약이 있을 것으로 예상되어 이러한 결과가 나타된 것으로 보인다. 지상 층수 뿐만 아니라 지하 층수도 화재 피해에 미치는 영향이 다를 것으로 예상되어, 지하 층수에 따른 평균 사망자 수, 부상자 수, 인명피해를 함께 분석하였다. 분석 결과, 지하 층수와 화재 피해 정도는 큰 연관성을 띄지 않는 것으로 파악되었다.



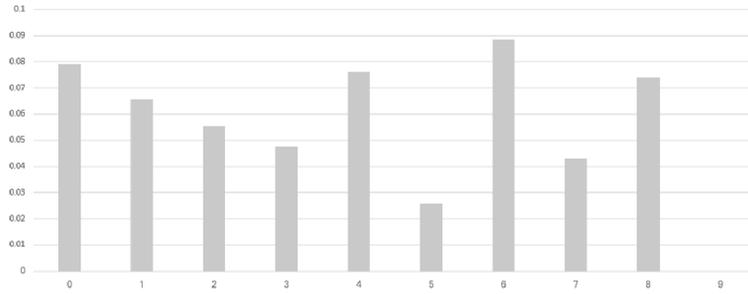
[그림 4-15] 지하 층수에 따른 평균 사망자 피해

출처 : 서울시 내부자료를 활용하여 연구진 작성



[그림 4-16] 지하 층수에 따른 평균 부상자 피해

출처 : 서울시 내부자료를 활용하여 연구진 작성



[그림 4-17] 지하 층수에 따른 평균 인명피해
출처 : 서울시 내부자료를 활용하여 연구진 작성

지하 층수에 따른 평균 사망자 수, 부상자 수, 인명피해를 함께 정리한 표는 아래와 같다. 분석 결과 현재 데이터를 바탕으로는 9층 이상의 지하 건물에서 화재로 인한 피해 발생이 없었던 것으로 나타났다.

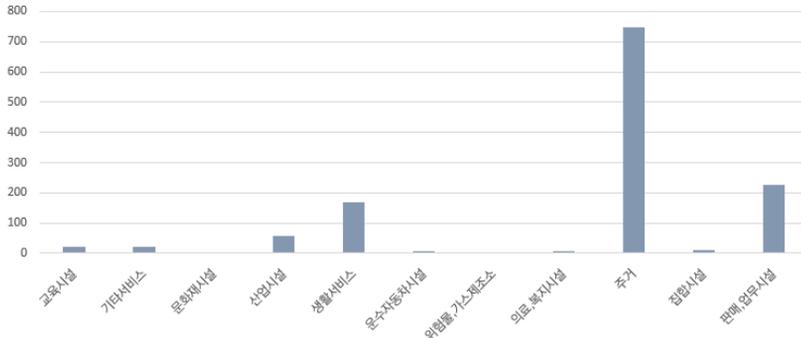
[표 4-6] 지하층수에 따른 평균 피해 정도

지하층수	평균 사망자수	평균 부상자수	평균 : 인명피해(명)소계
0	0.008042	0.070956	0.078997
1	0.008631	0.057049	0.06568
2	0.001761	0.053697	0.055458
3	0.003984	0.043825	0.047809
4	0.002924	0.073099	0.076023
5	0.005181	0.020725	0.025907
6	0	0.08871	0.08871
7	0	0.043011	0.043011
8	0	0.074074	0.074074
9	0	0	0

출처 : 서울시 내부자료를 활용하여 연구진 작성

④ 건축물 용도 및 사용승인 연도별 화재 피해

건축물 용도에 따른 화재 피해 정도를 파악하기 위해 용도별 화재 인명피해를 살펴 보았다. 건축물 용도로는 교육시설, 기타서비스, 문화재시설, 산업시설, 생활 서비스, 운수자동차시설, 위험물, 가스제조소, 의료 복지시설, 주거, 집합시설, 판매 업무시설 등이 있는데, 그 중에서도 주거, 판매 업무시설, 생활 서비스 시설 순으로 화재 발생 시 인명피해가 큰 것으로 나타났다.



[그림 4-18] 건축물 용도별 화재 인명피해 히스토그램

출처 : 서울시 내부자료를 활용하여 연구진 작성

[표 4-7] 건축물 유형별 따른 인명피해

건축물 유형	합계 : 인명피해 (명) 소계	건축물 유형	합계 : 인명피해 (명) 소계
교육시설	22	위험물, 가스제조소	0
기타서비스	23	의료, 복지시설	8
문화재시설	0	주거	749
산업시설	56	집합시설	11
생활서비스	168	판매, 업무시설	226
운수자동차시설	7	-	-

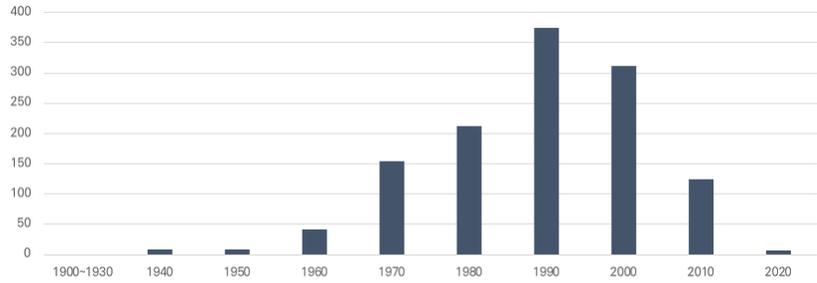
출처 : 서울시 내부자료를 활용하여 연구진 작성

⑤ 건축물 사용승인 연도별 화재 인명피해

건축물의 승인연도에 따라 건물의 노후도가 달라질 것으로 예상하여 건축물의 승인연도 별 화재 인명피해를 분석하였다. 1900년대~2020년대 까지 10년씩 년도를 구분하여 분석하였고, 1900~1930년대까지의 건축물에서 발생한 화재 인명피해는 없었기 때문에 단일 그룹으로 분석하였다.

주로 1990년대 준공된 건물에서 화재 인명피해가 많이 발생하는 것으로 나타났다. 오히려 사용승인연도가 더 오래된 건물에서는 적게 발생하였는데, 이는 재건축 연한과 관련이 있을 것으로 예상된다.

현행 재건축 연한은 30년이며, '도시 및 주거환경정비법(도정법)'에 따르면 준공 후 20년 이상~30년 이하의 건축물을 노후불량건축물로 관리할 수 있도록 하고 있다. 지은 지 30년이 넘는 건축물은 재건축 대상이 되기 때문에 1990년대 이전의 건물들은 많이 재건축되거나 사라진 건물들이 많을 것이므로, 따라서 대부분 현존하는 오래된 건물들 중에서는 1990년대 건물이 많을 것으로 추측된다. 따라서 결과 또한 1990년대 건축물이 화재 인명피해가 정점인 것으로 나타났다.



[그림 4-19] 건축물 사용승인 연도별 화재 인명피해

출처 : 서울시 내부자료를 활용하여 연구진 작성

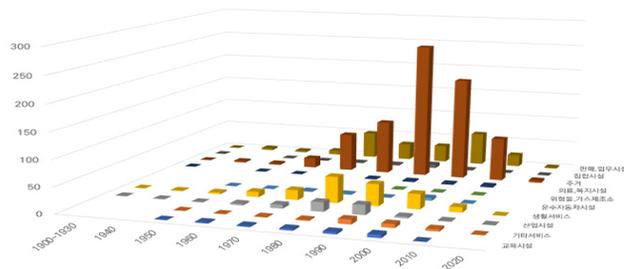
[표 4-8] 건축물 사용승인 연도별 인명피해

승인연도	합계 : 인명피해 (명) 소계	승인연도	합계 : 인명피해 (명) 소계
1900~1930	0	1980	212
1940	8	1990	374
1950	9	2000	311
1960	42	2010	124
1970	155	2020	6

출처 : 서울시 내부자료를 활용하여 연구진 작성

⑥ 건축물 사용승인 연도 및 시설 종류에 따른 화재 인명피해

건축물과 화재 피해와의 관계를 좀 더 다각적으로 분석하기 위해 건물 승인연도와 시설 종류 두가지의 측면에서 한번에 화재 인명피해를 분석하였다. 그 결과 생활서비스 시설의 경우는 1980년대 건물에서 인명피해가 높게 나타나는 모습을 보이고 있으며, 주거 시설에서는 1990년대, 2000년대가 월등히 높은 수치를 기록하고 있다. 판매, 업무시설의 경우는 오히려 1970, 1980년대, 2000년대에 다른년대보다 높은 인명피해를 보인다.



[그림 4-20] 건축물 사용승인 연도 및 시설종류에 따른 화재 인명피해

출처 : 서울시 내부자료를 활용하여 연구진 작성

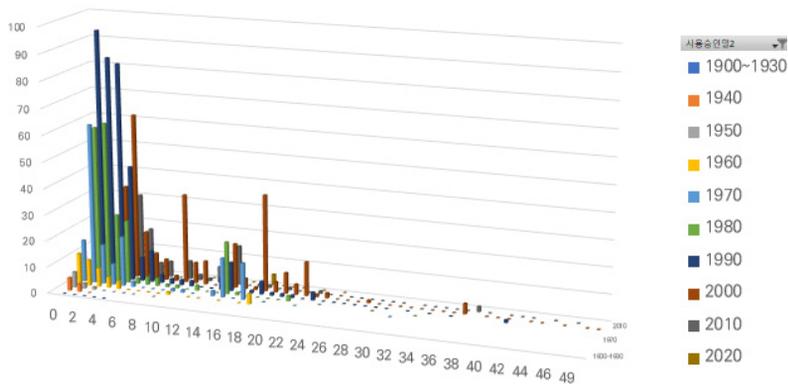
[표 4-9] 건축물 사용승인 연도 및 용도 별 화재 인명피해

구분	1940	1950	1960	1970	1980	1990	2000	2010	2020	총합계
교육시설	-	-	2	2	3	4	6	-	-	17
기타서비스	-	-		1	2	9	7	4	-	23
산업시설	-	-	3	7	18	19	3	1	-	51
생활서비스	1	3	10	19	50	42	29	10	-	164
운수자동차시설	-	-	0	1	1	0	5	0	-	7
의료, 복지시설	-	-	0	0	1	2	2	3	-	8
주거	3	3	19	73	103	258	195	83	4	741
집합시설	-	-	-	1	2	7	1	-	-	11
판매업무시설	4	3	8	51	32	33	63	23	2	219
총합계	8	9	42	155	212	374	311	124	6	1241

출처 : 서울시 내부자료를 활용하여 연구진 작성

⑦ 건물 지상층수와 건물 사용승인년도에 따른 화재 인명피해

건축물과 화재 피해와의 관계를 좀 더 다각적으로 분석하기 위해 건물 지상층수와 건물 승인년도, 두가지의 측면에서 화재 인명피해를 분석하였다. 그 결과 주로 1970년대, 1980년대, 1990년대, 2000년대 건축물의 화재 인명피해가 월등히 높았으며, 특히 지상 층수가 낮은 건물에서 화재로 인한 인명피해가 많이 발생하였는데, 이는 상대적으로 오래된 건물일수록, 지상 층수가 낮은 건물들의 개수가 많기 때문에 이러한 현상이 나타날 수 있다는 것 또한 생각해 볼 수 있다.



[그림 4-21] 건물 지상층수와 건물 승인년도에 따른 화재 인명피해

출처 : 서울시 내부자료를 활용하여 연구진 작성

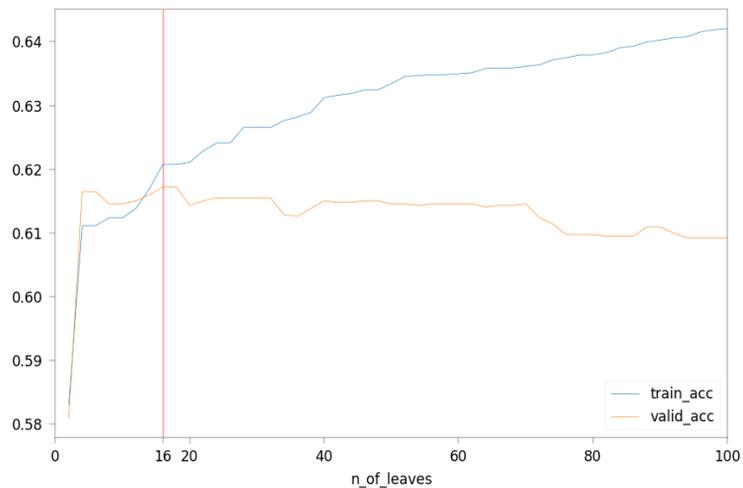
2. 기계학습 기반 화재 취약 건축물 예측

1) 의사결정나무 모델을 활용한 화재 취약 건축물 유형 추출

① 의사결정나무 모델 규모 결정

의사결정나무는 모델을 구성하는 노드의 수가 증가할수록 더 복잡하고 정확한 예측이 가능하지만, 지나치게 큰 모델은 학습에 사용한 훈련 데이터셋에 과적합 되는 경향이 있다. 본 연구의 분석에서도 모델 규모에 대한 제약조건 없이 학습시킨 결과 분류 유형의 수에 해당하는 잎 노드(leaf nodes)가 7,579개인 복잡한 구조의 모델이 도출되었는데, 훈련 데이터셋에 포함된 건축물 데이터의 크기가 32,892동임을 감안하면 각 유형에 평균 4.3개의 건축물을 포함하는 유형 분류를 수행하는 셈이다. 이 모델은 훈련 데이터셋에는 100% 정확도를 보이지만 검증 데이터셋에는 57.9%만 정확한 모형이 학습되는 현상을 보였다.

검증 데이터셋을 활용하여 의사결정나무 모델의 과적합을 방지하는 방법은 모델의 규모를 증가시켜가면서 검증 데이터셋에 대한 정확도를 측정하고, 정확도가 가장 높은 모델 규모를 선택하는 것이다. 본 연구에서 사용한 건축물 재난·재해 데이터에 대한 예측에서, 의사결정나무 모델은 잎 노드 16개 이후 검증 데이터셋에 대한 정확도가 오히려 감소하는 과적합 현상이 발생하였다. 이에 따라 화재 취약 건축물 예측 및 유형 추출에 적용하는 의사결정나무 모델의 규모를 잎 노드 16개로 설정하였다.



[그림 4-22] 의사결정나무 모델 규모에 따른 훈련 및 검증 데이터셋 예측 정확도

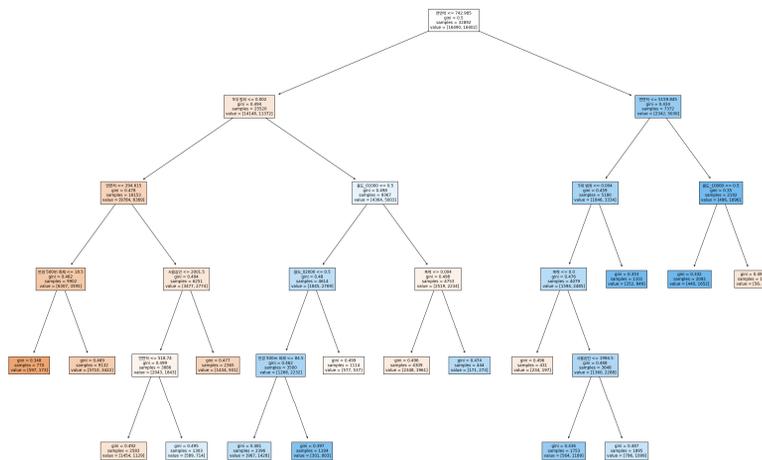
출처 : 연구진 작성

과적합 방지를 위한 모델 규모 결정 과정에 검증 데이터셋을 사용하였기 때문에 훈련 데이터셋에 대한 예측 정확도와 마찬가지로 검증 데이터셋에 대한 정확도도 모델의 실제 정확도를 보여줄 수 없다. 훈련 및 검증 데이터셋에 포함되지 않은 데이터에 대한 모델의 정확도를 편향 없이 측정하기 위해서는 시험 데이터셋에 대한 정확도를 사용하여야 한다.

② 화재 취약 건축물 유형 추출

앞에서 결정한 잎 노드 16개 규모 의사결정나무 모델을 훈련 데이터셋으로 학습시키고, 훈련 및 모형 규모 결정 과정에서 사용되지 않은 시험 데이터셋에 대한 예측을 수행한 결과, 의사결정나무 모델의 최종 정확도는 61.0%로 나타났다. 이는 훈련 및 검증 데이터셋에 대한 것과 유사한 수준으로, 학습 과정에서 과적합 현상이 나타나지 않았음을 확인하였다.

총 16개 잎 노드는 각각 다양한 조건에 따라 분류된 유형에 해당한다. 본 연구에서 적용한 의사결정나무 모델은 각 분기에서 유형을 두 가지로만 분류하는 이진 트리(binary tree)의 구조를 가진다. 건축물 재난·재해 데이터를 학습한 의사결정나무 모델에서 각 유형은 최소 3단계에서 최대 5단계의 분기를 거쳐 잎 노드에 도달하는데, 각 분기는 그 유형을 정의하는 한 가지 조건에 해당한다. 예를 들어, 5개 분기를 거쳐 잎 노드에 도달하는 유형은 서로 다른 5가지 조건으로 정의된다. 단, 여러 분기가 하나의 변수에 대한 서로 다른 조건을 나타낼 수 있기 때문에, 최종적인 유형 특성의 수가 분기의 수와 항상 같지는 않다.



[그림 4-23] 건축물 화재 발생 예측 의사결정나무 모델(파랑: 화재 발생률 높음, 주황: 낮음)

출처 : 연구진 작성

총 16개 읍 노드 중 8개 노드가 훈련 데이터셋 기준 전체 대비 화재 발생률이 높은 조건을 보였는데, 그 중에서 화재 발생률이 높은 상위 5개 유형을 선택하여 화재 취약 건축물 유형을 추출하였다. 이들 유형에는 총 12.9%의 건축물이 포함되며, 최대 12.0%의 화재 발생률을 보여 전체 건축물(3.8%) 대비 최대 3.2배까지 취약한 건축물 유형이 추출되었다.

의사결정나무 모델 구조를 통하여 화재 취약 건축물 유형의 특성을 분석한 결과는 [표 4-10]과 같다. 검토 결과, 화재 취약 건축물 유형 모두에서 연면적, 용도 등 건축물 특성과의 연관성을 보여, 건축물 안전 예측에 건축물 특성에 대한 고려가 필요함이 입증되었다. 또한, 5개 유형 중 4개 유형이 5대 범죄 발생률 또는 폭력 범죄 발생률과 연관이 있는 것으로 나타나 화재 및 범죄 데이터 연계 필요성을 검증하였다. 마지막으로 1개 유형에서 반경 500m 이내 화재 발생 빈도와 연관성을 보여 건축물 재난·재해의 공간적 상관성에 대한 고려가 필요함을 확인하였다.

5개 취약 건축물 유형 중 화재 발생률이 가장 높은 유형은 연면적 5,160㎡ 이상 건축물 중 교육연구시설을 제외한 유형으로, 이 유형에 해당하는 약 1.7만 동의 건축물은 최근 5년 화재 발생률이 12.0%에 달했다. 가장 다양한 데이터로 정의된 취약 건축물 유형은 연면적 743㎡ 미만, 5대 범죄 발생률 0.002건/㎡ 이상, 단독주택(대분류) 및 공동주택 제외, 반경 500m 화재 발생 85건 이상 등의 조건에 해당하는 약 1.3만 동의 건축물로, 건축물 속성, 인근지역 화재 및 범죄 발생률 등을 모두 사용하여 최근 5년 화재 발생률 8.1%인 건축물 유형을 정의하였다.

[표 4-10] 재난·재해 취약 건축물 유형별 화재 발생률 비교

특성	건축물 수	화재 발생률
연면적 5,160㎡ 이상, 교육연구시설 제외	17,345	12.0%
연면적 743㎡ 이상 5,160㎡ 미만, 5대 범죄 발생률 0.004건/㎡ 이상	10,720	10.7%
연면적 743㎡ 미만, 5대 범죄 발생률 0.002건/㎡ 이상, 단독주택(대분류) 및 공동주택 제외, 반경 500m 화재 발생 85건 이상	12,566	8.1%
연면적 743㎡ 이상 5,160㎡ 미만, 5대 범죄 발생률 0.004건/㎡ 미만, 폭력 범죄 발생 있음, 사용승인 1994년 이전	21,547	6.9%
연면적 743㎡ 미만, 단독주택(대분류), 폭력 범죄 발생률 0.004건/㎡ 이상	7,288	4.8%
전체 건축물	540,111	3.8%

출처 : 연구진 작성

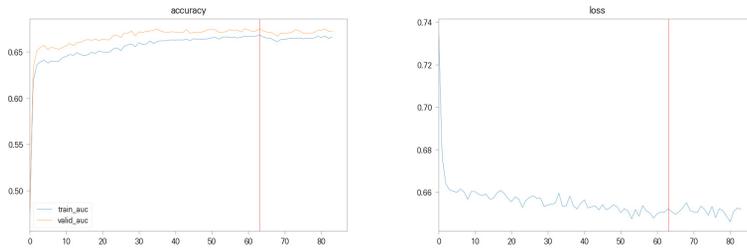
2) TabNet 모델을 활용한 화재 위험 영향 요인 추출

① 딥러닝 학습 진행 및 최적 모델 선택

딥러닝 모델은 학습을 반복적으로 진행할수록 정확도가 높아지지만, 일정 수준 이상이 되면 과적합 현상이 발생하게 된다. 딥러닝 모델의 과적합을 방지하는 방법 중 하나는 검증 데이터셋에 대한 정확도를 주기적으로 측정하여 학습이 진행될수록 정확도는 오히려 낮아지는 경우 학습을 조기에 중단(early stopping)하는 방법이 있다.

TabNet 모델에 건축물 재난·재해 데이터를 적용하여 딥러닝 학습을 진행한 결과, 예측모형은 매우 빠르게 60%대 정확도에 도달하였으나 이후 매우 더딘 학습 속도를 보이며 63기(epoch) 이후 검증 데이터셋에 대한 정확도는 오히려 감소하면서 조기 중단되었다. 이에 따라 화재 취약 건축물 예측 및 화재 위험 영향 요인 추출에 적용하는 딥러닝 모델에 검증 데이터셋에 대한 정확도가 가장 높았던 최적 모델을 사용하였다.

해당 모델의 검증 데이터셋에 대한 정확도는 67.5%로 나타났으나, 학습 및 조기 중단 판단에 사용되지 않은 시험 데이터셋에 대하여 평가한 최종 정확도는 그보다 약간 낮은 65.8%로 나타났다. 이는 의사결정나무 모형의 61.0%에 비하여 약간 높은 수준으로, 정확도와 해석 기능성과의 상충 관계를 고려하여 기계학습 모델을 선택할 필요가 있음을 확인하였다.



[그림 4-24] TabNet 모형 학습 과정

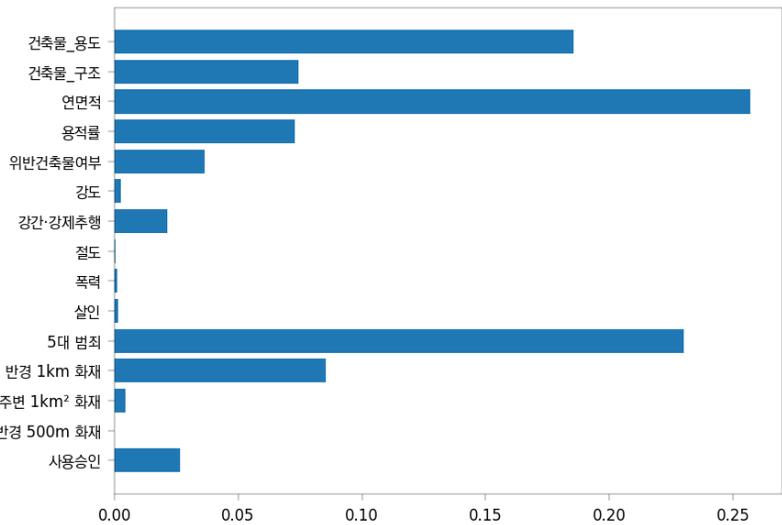
출처 : 연구진 작성

② 화재 위험 영향 요인 추출

TabNet 모델은 모델이 학습한 규칙에 대한 자체적인 해석 방법론으로, 모델이 수행하는 분류 또는 회귀에 각 변수가 미치는 영향을 변수 중요도로 나타내는 방법론을 포함한다. 건축물 재난·재해 데이터를 학습한 최적 TabNet 모델에서 합계를 1로 하여 각 변수의 중요도를 나타낸 결과는 [그림 4-25]와 같다.

화재 발생 여부의 예측에 가장 중요하게 사용된 변수는 건축물 연면적, 인근지역 5대 범죄 발생률, 건축물 용도 순으로 나타났다. 인근지역 건축물 화재 발생의 경우는 반경 1km 발생 건수가 그 다음으로 중요하게 나타났으나, 그 중요도는 직전 상위인 건축물 용도의 절반 이하로 제한적이었다.

반대로 가장 중요성이 낮은 변수의 경우, 인근지역 반경 3개 중 2개 화재 발생 건수와 5대 범죄 중 4개 범죄 발생률 등이 화재 예측에 거의 사용되지 않은 것으로 나타났다. 유사한 변수 중 일부는 높은 중요도를 보이지만 나머지는 중요도가 거의 없는 것은 TabNet 모델이 유사한 종류의 정보를 포함한 상관성이 높은 여러 변수 중 가장 적합한 것 하나를 스스로 선택하여 사용할 수 있다는 의미로, 이를 통하여 데이터 분석에서 중요한 단계 중 하나인 변수 선택을 자동으로 수행할 수 있음을 확인하였다.



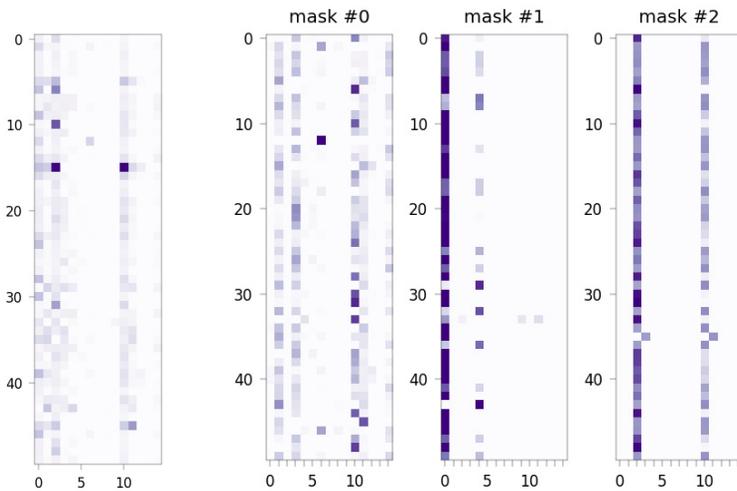
[그림 4-25] 건축물 화재 예측에서 TabNet 모델의 변수 중요도

출처 : 연구진 작성

TabNet 모델은 전체 데이터에 대한 중요도뿐만 아니라, 모델이 각 사례별로 종합적으로, 또 각 단계마다 사용한 변수와 그 중요도를 추출하여 분석할 수 있다. [그림 4-26]은 일부 사례에 대하여 종합 및 단계별 변수 중요도를 색으로 나타낸 것이다. 가장 왼쪽 행렬은 각 사례에 대한 종합 변수 중요도를 나타내며, 오른쪽은 TabNet 모델의 각 단계마다 이루어지는 변수 선택에 따른 중요도를 나타낸다. 각 행렬에서, 각 행은 건축물 한 동에 대한 화재 발생 여부 예측에 사용된 각 변수의 중요도를 나타내며, 각 열은 건축물 용도부터 사용승인까지 총 15개 변수를 0부터 세어 0~14의 값으로 나타낸다.

가장 왼쪽 행렬에서는 건축물 용도(0번 열), 연면적(2번 열), 5대 범죄 발생률(10번 열) 등이 짙은 색의 열로 나타나는데, 많은 사례에서 중요하게 사용되었다고 해석할 수 있다. 각 단계별 중요도에서도 0부터 세어 두 번째(1번) 단계에서는 건축물 용도(0)가, 세 번째(2번) 단계에서는 연면적(2)과 5대 범죄 발생률(10)이 거의 대부분의 사례에서 가장 중요한 변수로 사용된 것을 확인할 수 있다. 이러한 결과는 전체 데이터에 대한 종합 변수 중요도 분석 결과와 일치한다.

한편, 가장 중요하게 사용된 3개 변수 외에도 다양한 변수가 화재 예측에 사용되고 있으며, 각 사례별로 다양하게 적용되고 있음을 확인할 수 있다. 첫 번째(0번) 단계에서는 건축물 구조(1)와 용적률(3) 변수도 중요하게 사용되며, 반경 1km 화재 발생 건수(11)와 사용승인 연도(14) 등 변수도 함께 사용되는 것을 볼 수 있다. 이러한 변수들은 최종 예측 결과에 직접 영향을 주지는 않지만, 다음 단계의 변수 선택 등에 영향을 미쳐 결과적으로 간접적인 영향을 미친다.



[그림 4-26] TabNet 모델의 각 사례별 종합 및 단계별 변수 중요도 (짙음: 중요도 높음)

출처 : 연구진 작성

각 사례별로 변수 선택 및 판단 과정에 따라 다양한 중요도가 해석 결과로 도출될 수 있다는 것은, 의사결정나무 모델의 경우와 유사하게 각 건축물에 대한 예측 결과가 도출된 원인을 분석하고 유형화하기 위한 기초자료를 확보할 수 있다는 것을 의미한다. 최근 딥러닝 연구는 이렇게 딥러닝 모델의 해석 가능성을 높이는 방법론에 대한 다양한 접근을 시도하고 있다. 본 연구의 적용 결과에서도 이러한 방법론의 가능성을 확인할 수 있었으며, 앞으로 학습 결과에 대한 해석이 필요한 연

구 및 활용의 경우에도 딥러닝 모델을 사용할 수 있는 가능성이 높아질 것을 기대할 수 있다.

3) 화재 취약 건축물 예측 및 시각화

의사결정나무 모델과 TabNet 모델 모두 다른 기계학습 모델과 마찬가지로 전체 데이터셋에 대한 예측을 수행할 수 있으며, 따라서 개별 건축물의 화재 발생 여부 예측 결과가 도출될 수 있다. 앞에서 논의한 바와 같이, 화재가 발생하지 않은 건축물에 대한 잘못된 화재 발생 예측의 경우 화재 발생 건축물과 화재 발생 연관 특성을 공유하는 유사한 건축물, 즉 화재 발생 위험성이 높은 건축물로 해석할 수 있다.

이러한 예측 결과를 'GIS건물통합정보'의 건축물 위치정보와 연계하여 시각화한 결과, 위치정보가 배제된 기계학습 분석 결과에 대한 해석에서 파악하기 어려웠던 패턴을 시각적으로 확인할 수 있었다. 도로변 건축물의 경우 상대적으로 높은 용적률과 넓은 연면적이라는 속성을 지니고 있으므로, 해당 변수를 바탕으로 판단한 결과 화재 발생 위험성이 높은 건축물이 도로변에 위치한 것으로 나타났다. 그 외에도 지역·지구와 연관성이 낮은 건축물 속성에 따라 개별 건축물 단위로 주변 건축물과 다른 결과를 보이는 경우가 있는데, 위치정보에 따른 시각화를 통하여 다양한 건축물의 화재 발생 위험성을 쉽게 파악할 수 있었다.



[그림 4-27] 건축물 화재 발생가능성 예측 결과(발생: 빨간색, 미발생: 파란색)

출처: 연구진 작성

3. 소결 및 시사점

본 장에서는 화재 발생 데이터와 건축물 속성정보를 결합하여 건축물 화재의 특성에 대한 기술통계를 수행하였다. 또한 기계학습 기반 건축물 재난·재해 예측 모델을 화재 취약 건축물 예측에 적용하여, 건축물 속성과 인근 지역의 화재 및 범죄 발생 등이 공간적으로 연계된 데이터를 기반으로 건축물 화재 발생 여부를 예측하는 모델을 학습시키고, 이를 통해 화재 취약 건축물, 화재 취약 건축물 유형, 건축물 화재 위험 영향 요인 등을 추출하였다.

1) 건축물 화재 데이터 기술통계

본 장에서는 화재 건축물 통합데이터를 활용하여 사고와 건축물 속성에 대한 기술통계 분석을 진행하였다.

서울시 건축물 화재 발생 건수 통계를 분석한 결과 건축물 용도별로는 공동주택 및 단독주택에서의 화재 발생 건수가 높았으며, 다음으로 근린생활시설에서의 화재 발생 건수가 많았다. 건축물 연면적별 화재 통계에서는 200㎡초과 500㎡이하의 건축물에서 화재 발생 건수가 많았으며, 층수별 화재 통계에서는 5층 이하의 저층에서 화재발생 건수가 많았다. 건축물 노후도별 화재발생 통계를 살펴보면 일반적으로 노후화가 진행될수록 화재발생 건수가 많음을 알 수 있다.

화재 건축물 통합데이터를 통해서 서울시 구/동별 화재피해와, 인구나 면적 대비 화재피해, 지상 층수 및 지하 층수에 따른 화재피해, 건축물 용도별 화재피해, 건축물 사용승인 연도별 화재피해의 관계를 도출하였으며, 더 나아가 건축물 사용승인 연도 및 용도에 따른 화재피해, 지상층수 및 건축물 사용승인 연도에 따른 화재피해까지 살펴보았다.

화재 발생과 건축물의 특성을 함께 비교하기 위해 지상 층수 및 지하 층수에 따른 화재 피해 정도를 비교한 결과 평균 사망자 수, 부상자 수, 인명피해, 재산피해 4가지 항목에 대해서 대체로 지상 층수가 높은 건물일수록 더 많은 피해가 있는 것을 확인할 수 있었다.

지상 층수 뿐만 아니라 지하 층수도 화재 피해에 미치는 영향이 다를 것으로 예상되어, 지하 층수에 따른 평균 사망자 수, 부상자 수, 인명피해를 함께 분석하였지만 지하층수와 화재 피해 정도는 큰 연관성을 띄지 않는 것으로 나타났다.

건축물 용도별 화재 피해는 주거, 판매 업무시설, 생활 서비스 시설 순으로 화재 발생 시 인명피해가 큰 것으로 나타났다. 또한 건축물의 사용승인연도 별 화재 인

명피해를 분석하였다. 주로 1990년대 건물에서 화재 인명피해가 많이 발생하였는데, 이는 지은 지 30년이 넘은 건축물은 재건축 대상이 되기 때문에 대부분 현존하는 오래된 건물들 중에서는 1990년대 건물이 많을 것이므로, 1990년대 건축물이 대체로 화재 인명피해가 많은 것으로 나타났다.

건축물과 화재 피해와의 관계를 좀 더 다각적으로 분석하기 위해 건물 승인년도와 시설 종류, 지상층수와 건물 승인년도 측면에서 화재 인명피해를 분석하였다. 그 결과 생활서비스 시설의 경우는 1980년대 건물에서 인명피해가 높게 나타나는 모습을 보이고 있으며, 주거 시설에서는 1990년대, 2000년대가 월등히 높은 수치를 기록하는 등 용도와 노후도에 따라 인명피해가 조금씩 다른 양상을 보이는 것을 확인 할 수 있었다. 지상층수와 노후도와의 관계에서는 1970년대, 1980년대, 1990년대, 2000년대 건축물, 그리고 저층 건축물에서 화재 인명피해가 전반적으로 많이 발생하는 것을 확인하였다.

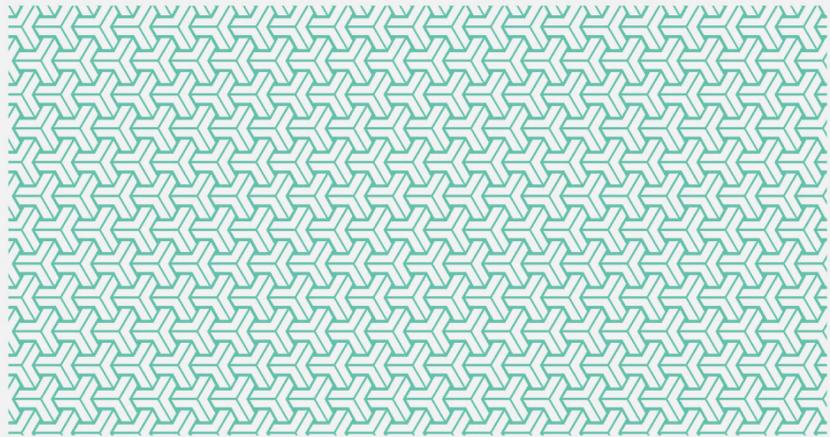
2) 기계학습 기반 화재 취약 건축물 예측

빅데이터 분석에서는 의사결정나무 모델과 딥러닝 모델 중 TabNet 모델을 활용하여 화재 취약 건축물을 예측하고, 기계학습 모델에 대한 분석을 통하여 화재 취약 건축물 유형 및 화재 위험 영향 요인을 추출하였다. 이러한 과정을 거쳐 예측 모델을 검증하고 화재 취약 건축물의 특성을 해석하였다. 의사결정나무 모델을 통한 분석 결과 연면적 5,160m² 이상인 건축물 중 교육연구시설을 제외한 유형에서의 화재 발생률이 12.0%로 가장 높았다.

TabNet 모델에서 도출된 화재 위험 영향 요인 추출을 통하여, 딥러닝 모델에 대한 해석을 통한 예측 모델 활용의 가능성을 검증하였다. 분석 결과 검증 데이터셋에 대한 정확도는 67.5%로 나타났으며, 학습 및 조기 중단 판단에 사용하지 않은 시험 데이터셋에 대한 정확도는 65.8%로 나타났다. 화재 발생과 연관성이 높은 변수는 연면적, 5대범죄 발생건수, 건축물 용도 순으로 나타났다.

본 연구에서 제안한 기계학습 기반 건축물 재난·재해 예측 모델은 개별 건축물 단위로 화재 취약성 예측, 해당 취약 유형, 취약 판단이 도출된 원인 등을 파악할 수 있는 방법론이다. 분석의 마지막 단계로서 화재 취약성 예측 결과를 공간정보 기반으로 시각화하여, GIS 시각화 기술을 적용한 활용 가능성을 검토하였다.

제5장 결론



1. 연구의 요약
2. 향후 연구과제 및 정책적 시사점

1. 연구의 요약

본 연구는 재난·재해 확산 등 사회변화에 따라 건축물의 대형 안전사고가 증대하고, 사회발전에 의한 중대형 건축물의 증대로 안전사고 발생 시 대형 인명 및 재산피해 우려가 높아지고 있어, 빅데이터 기반 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델을 개발하였다. 이를 위하여 건축·도시공간의 데이터 및 유관정보의 연계·활용 방안을 도출하고, 선행연구와 사례 분석을 통하여 방법론과 모델을 개발하였다. 또한 개발한 모델을 실제 대상지에 적용하여 건축물 화재발생 예측하여 모델의 타당성을 검증하였다.

□ 재난·재해 빅데이터 시스템 운영 및 연구 동향

국가는 재난 및 재해 예방·관리·복구를 위해 다양한 위기관리 시스템을 구축하여 운영하고 있으며, 관련 시스템은 다양한 재난·재해 데이터를 기반으로 운영되고 있다. 국가재난관리정보시스템은 4가지 시스템을 운영하고 있으며, 풍수해, 지진, 침수, 화재 등의 재난에 대한 관리·전파·대피·대응 등에 대한 정보를 제공하고 있다. 안전정보통합공계시스템은 건축물과 연계하여 안전관리 정보를 제공하고 있고, 특히 재난·치안·교통·보건·생활·시설에 대한 생활안전 지도를 제공하고 있다. 이외에도 화재와 범죄는 국가화재정보시스템 및 지리적 프로파일링 시스템을 통해 관련 정보 및 발생 위험지역 예측을 수행하고 있다.

국외의 경우 화재 및 풍수해 등에 대한 모니터링 시스템을 운영하고 있으며, 다양한 변수를 고려하여 예측 프로그램을 제공하고 있다. 또한 정책적 활용을 위해 재난·재해에 대한 지수·지표를 제공함으로써 예방·대응·확산 방지를 위한 정책적 판단을 지원하고 있다.

빅데이터 분석 방법론을 적용한 연구 동향을 살펴보면 예측을 위해 활용되어야 하는 데이터 유형 및 목록과 기존 데이터를 활용하는 빅데이터 연구방법론에 초점이 맞춰져 있다. 관련 연구 고찰을 통한 시사점은 재난·재해 위험지역 예측 모델 구축을 위해서는 재난·재해 대응 시나리오가 중요하며, 시나리오를 면밀히 구현하기 위한 분석 모델의 선정에 대한 고려가 있어야 한다는 것이다. 또한 가능한 다양한 변수들을 투입함으로써 기계학습의 정교성을 높일 수 있어야 보다 정확한 예측 모델 구축이 가능함을 시사하고 있다.

□ 재난·재해 위험지역 예측 방법론과 건축물 화재 예측 모델 개발

재난·재해 위험지역 예측 방법론은 크게 대상지 선정-원시데이터 수집 및 정제-

변수구축 및 데이터 전처리-빅데이터 분석 수행 순으로 이루어진다. 우선 재난·재해 위험지역 예측을 위한 대상지 선정에 있어 고려해야 할 사항은 예측하고자 하는 재난·재해 사고 발생 건수가 빅데이터 분석을 적용할 만큼 많아야 한다. 또한 재난·재해와 연관성 있는 많은 변수들이 고려되어야 하기에 다양한 인문사회학적 요소 및 물리적 요소를 갖춘 곳으로 선정하여야 한다. 이에 본 연구에서는 서울시를 그 대상으로 하였다. 서울시는 다수의 화재 사고가 발생하였으며, 대도시인 만큼 다양한 인구사회학적 요소를 가지고 있고, 건축물의 분포도 다양하기 때문이다.

재난·재해 데이터 수집 및 정제에서는 빅데이터 분석이 가능한 충분한 양의 데이터를 확보하여야 한다. 이에 대한 명확한 기준은 정립되지 않았지만 일반적으로 데이터의 크기로 인해 일반적인 통계적 기법 적용이 어려울 정도의 크기를 말한다. 또한 위치정보를 특정할 수 있는 식별부호가 있어야 한다. 재난·재해 연관 변수 데이터는 예측하고자 하는 재난·재해와 연관성이 높을 것으로 가정한 데이터이며, 재난·재해 데이터와 연계 가능하여야 한다. 본 연구에서는 서울시 화재 위험지역 예측모델을 구축하는 것이 목적이기 때문에 최근 5년간 화재발생 데이터를 활용하였으며, 주요 변수로는 건축행정데이터 및 범죄데이터를 활용하였다.

변수 구축 및 데이터 전처리 절차에서는 우선적으로 빅데이터 분석 방법론 적용을 위한 정제된 변수의 가공 및 변수를 확정하여야 한다. 이후 분석을 위한 기계학습 모델을 선정하여야 하는데, 활용하고자 하는 재난·재해 데이터 및 변수 분석에 적합한지 여부와 해석가능성과 예측 성능 측면에서 장단점이 보완될 수 있는 모델을 선정하여야 한다. 데이터 전처리 과정에서는 학습용 및 검증 데이터를 구축하고 오류 발생 가능성을 보완할 수 있도록 한다. 본 연구에서는 기계학습 모델로 의사결정나무와 TabNet을 선정하였다. 모델 선정 후 데이터 전처리 과정에서 다양한 방식으로 균형있는 분류범주, 변수 이진화, 결측값 대체 과정을 진행하였다.

□ 빅데이터 기반 건축물 화재 예측 모델 적용 결과

개발한 건축물 안전사고 예측 모델을 타당성을 검토하기 위하여 건축물 화재 발생 위험을 측정하였다. 건축물 화재 발생 기술통계를 살펴보면 건축물 용도의 경우 공동주택 및 단독주택, 연면적의 경우 200㎡초과 500㎡이하에서 화재발생 건수가 많았으며, 건축물이 저층이고, 노후화가 진행될수록 화재 발생 건수가 많았다. 고층 건물일수록 화재피해 규모가 컸음을 알 수 있다. 이는 고층 건물의 경우 화재 발생 시 진압에 소요되는 시간 및 제약이 있을 것으로 추정된다. 건축물 용

도별 화재 피해는 판매 업무시설, 생활 서비스 시설 순으로 나타났으며, 노후화된 건축물의 경우 화재 인명 피해가 큰 것으로 나타났다. 승인년도와 용도를 교차해서 분석한 결과 30년 이상 건축물은 생활서비스 시설에서, 20~30년 된 건축물은 주거시설에서 화재 피해가 큰 것으로 나타났다.

건축물 화재 위험지역 예측 모델 중 의사결정나무를 통해 건축물 화재를 분석한 결과 교육연구시설을 제외한 연면적 5,160㎡ 이상의 화재 발생률이 12%로 가장 높았으며, 연면적 743㎡ 이상 5,160㎡ 미만 및 5대범죄 발생률 0.004건/㎡에서 10.7%의 화재 발생률을 나타냈다.

딥러닝 모델인 TabNet을 통한 화재위험도를 분석한 결과 검증 데이터셋에 대한 정확도는 67.5%로 나타났으며, 학습 및 조기 중단 판단에 사용하지 않은 시험 데이터셋에 대한 정확도는 65.8%로 나타났다. 화재 위험 영향 요인 추출 결과 건축물 속성 중에서는 연면적이 가장 연관성이 높게 나타났으며, 용도가 다음으로 연관성이 높게 나타났다. 위반건축물 여부 및 사용승인(노후도)은 상대적으로 연관성이 낮게 나타났다. 범죄의 경우 5대범죄 건수가 연관성이 높게 나타난 반면, 개별 범죄건수는 연관성이 낮게 나타났다. 거리변수의 경우 화재발생 지역 반경 1km가 주변 1km² 및 반경 500m에 비해 연관성이 높게 나타났다.

2. 향후 연구과제 및 정책적 시사점

빅데이터 기반 건축물 화재 예측 모델 개발 연구는 서울시 화재발생 건축물 데이터를 기반으로 빅데이터 분석 방법론을 적용하여 위험지역 건축물을 예측하고, 이를 수행하는 재난·재해 위험지역 예측 방법론을 구축하였다. 다만 본 연구는 데이터 구축 및 예측모델의 정확성 측면에서 한계가 존재한다.

첫째, 서울시 건축물 화재발생 데이터와 건축행정정보시스템을 연계하는데 있어 각 데이터의 오류로 인해 모든 건축물에 대한 화재발생 정보를 연계할 수 없었다는 점이다. 2017~2021년 5년간 서울시 화재발생 데이터 21,055건 중 건축행정데이터와 매칭된 데이터는 18,814건으로 전체의 89.4%가 연계되었다. 연계 시 오류는 건축물 대장이 없거나 주소가 잘못 기입되어 있는 경우, 또는 화재 데이터의 오류 등이다.

둘째, 의사결정나무 및 TabNet 모델을 적용한 예측정확도는 의사결정 모형이 61%, TabNet 모델이 65.8%로 나타나 예측정확도가 상대적으로 낮다는 것이다. 이는 여러 요인이 있을 수 있으나, 본 연구에서 적용한 변수의 유형이 적기 때문

인 것으로 보인다. 또한 화재발생 건수의 경우 최근 5년치를 적용하였음에도 불구하고, 절대적인 데이터의 수가 적기 때문인 것으로 추정할 수 있다.

이에 본 연구를 기반으로 한 향후 연구과제 및 정책적 시사점은 다음과 같다.

본 연구의 목적은 빅데이터 분석 방법론에 기반한 건축물 화재 예측 모델의 개발에 있다. 이에 안전사고 예측 모델의 정확성을 높이기 위해 다양한 변수를 활용할 필요가 있다. 건축물의 속성정보 이외에도 침수 및 붕괴와 같은 건축물에서 발생한 다양한 재난·재해 발생 데이터 및 피해규모 데이터를 연계하는 것이 필요하다. 또한 재난·재해 데이터 이외에도 발생 지역의 인구유형, 경제적 상황, 취약가구 현황 등 인문·사회학적 데이터에 대한 변수를 구축하여 적용하는 것이 필요하다. 이를 통해 건축물 화재 예측 모델을 통한 예측 정확도를 높일 수 있을 것으로 기대된다.

또한 건축물 단위에서의 재난·재해 위험지역 예측에서 나아가 블록 등 공간단위의 위험지역 예측모델을 개발하는 방향으로 나아갈 필요가 있다. 건축물의 안전 점검 및 재난·재해 대응에 있어 특정 건축물에 대한 안전사고 예측 보다는 공간단위의 재난·재해 피해 위험지역 예측이 국가 및 지자체의 예방정책에 도움이 될 것으로 판단된다.

정책적으로 가장 시급하다고 보이는 것은 건축행정데이터 및 재난·재해데이터의 연계시 높은 수준의 무결성을 달성하는 것이다. 현행 건축행정데이터는 건축물 고유번호(PK번호)에 의해 데이터가 형성되어 있는데, 향후 안전사고 및 재난·재해 원시데이터와의 연계를 위해서는 건물ID 기반으로 연계되어야 할 것이다. 건물ID는 기존 공동고유번호 체계를 활용하되, 적용 시점을 건축물 인허가시기로 확대하여 건축물 생애기반의 관리 효율성 및 활용성을 제고하기 위해 구축되고 있는 실정이다. 이에 향후 건물ID 기반 데이터담 구축을 통해 정부가 보유하고 있는 공공데이터를 건축물 단위로 융합하여 가공 및 제공함으로써 효율과 효과를 동시에 제고할 수 있을 것으로 기대하고 있다. 이를 활용하여 빅데이터 분석방법론을 적용한 건축물 안전사고 및 재난·재해 위험지역 예측 정확도는 높아질 것으로 기대된다.

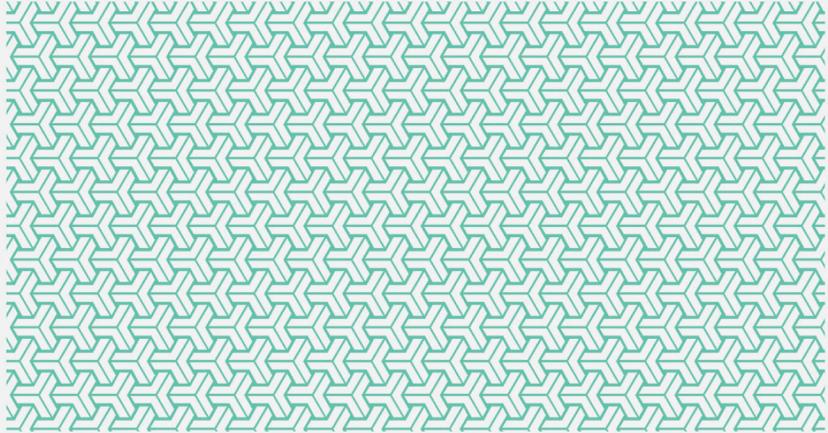
건물ID와 재난·재해데이터의 연계를 통한 안전사고 위험지역 예측을 통해 소방훈련 및 건축물 관리 등을 수행함에 있어 효율적으로 대체할 수 있을 것이다. 현행 소방훈련 대상 건축물은 「화재예방, 소방시설 설치·유지 및 안전관리에 관한 법률 시행령」 제22조의 규정에 의해 소방안전관리자를 두어야 하는 특정소방대상물 중 상시 근무 또는 거주인원이 10인 이상인 소방안전관리대상물로 규정되

어 있다. 건축 및 공간단위의 화재위험지역 예측이 가능하다면, 소방훈련의 대상 범위를 건물 단위에 더해 공간단위로 확대할 수 있을 것으로 보인다. 특히 화재 예측 비율이 높은 지역의 경우 좁은 골목길 및 대형소방차의 접근이 어려운 지역 일 경우 소방훈련을 통해 선제적으로 소방동선 파악을 할 수 있어 화재 발생시 피해를 줄일 수 있으며, 소방훈련을 통한 화재위험 요소를 파악하여 화재발생을 미연에 방지할 수 있을 것이다.

건축물 관리에 있어서도 건축물 화재 예측 모델은 유용할 것이다. 현행 건축물에서 수행되고 있는 소방점검은 주로 화재발생 시 피해최소화를 위한 시설점검을 중심으로 이루어지고 있다. 빅데이터 분석 방법론에 기반한 화재위험 예측 모델의 고도화 될 경우 특정 유형의 건축물에서의 발화요인에 대해 예측 가능할 것으로 기대된다. 이를 통해 현재 관련 제도에 따라 수행되고 있는 일정 규모와 용도에 따른 화재 발생 피해 최소화를 위한 안전점검뿐만 아니라, 위험이 예측되는 건축물을 대상으로 사전에 전기, 기계, 화학적 요인에 대한 점검을 수행하여, 화재 발생의 피해를 최소화하는 것이 아니라 예방을 통하여 화재 발생 자체를 줄일 수 있다.

참고문헌

References



□ 문헌자료

- 건축공간연구원(2021). 2021년 건축물관리지원센터 업무 위탁 보고서.
- 구인혁, 윤용기, 김동준, 김봉찬, 권영진(2016). 방재도시 구축을 위한 도시화재 위험성 평가 시스템의 구축 및 적용성 검토에 관한 연구. 한국방재학회 논문집, 16(6), 137-143.
- 국토교통부 보도자료 (2021.03.09.). 전국 건축물 총 7,275,266동/39억 6천만㎡.do, (검색일 : 2022.02.24.)
- 권설아, 이재은, 반응운, 이해진, 유성은, 유현정(2018). 다중이용시설의 화재 취약성 극복과 안전대책 : 밀양 화재참사와 제천 화재참사의 비교 분석. Crisisonomy, 14(5), 149-167.
- 김도형, 조병완(2021). 빅데이터 분석을 활용한 초기 정보 기반 화재현장 위험도 예측 모델 개발 연구. 한국재난정보학회논문집, 17(2), 245-253.
- 김보형(2020), 2020년 주요 안전사고 총정리, 안전저널(2020.12.18.)
- 김보형(2021), 2021년 주요 안전사고·이슈 총정리, 안전저널(2021.12.30.)
- 김용문, 이태식(2019). 서울시 재난 사례 QRE 평가도구를 활용한 재난 위험도 평가. 한국방재안전학회 논문집, 12(1), 11-21.
- 김영민, 황미녕, 김태홍, 정창후, 정도현(2015). 자연재해 분석을 위한 빅데이터 마이닝 기술. 한국데이터정보과학회지, 26(5), 1105-1115.
- 김태형(2015). 수요자 중심의 Smart City 구축을 위한 빅데이터 활용 방안: 재난안전 분야를 중심으로. 한국지역정보과학회지, 18(2), 47-68.
- 나유경, 박진홍, 최진무(2019). 빅데이터를 이용한 복합재난의 발생 지역 및 유형 분석. 국토지리학회지, 53(4), 375-384.
- 박무중, 박민규, 송영석(2011). 지역별 재난관리를 위한 위험인자의 공간분포 해석 : 1.재해위험성 및 노출성 인자. 한국방재학회논문집, 11(4). 189-199.
- 박민규, 박무중(2011). 지역별 재난관리를 위한 위험인자의 공간분포 해석 : 1.취약성. 한국방재학회논문집, 11(4). 201-209.
- 박유나, 손동필, 현태환(2020), 한국형 범죄예방환경설계를 위한 장소프로파일링 기법 개발 연구. 건축공간연구원.
- 박은석, 민세홍(2019). 빅데이터 활용을 위한 화재인자 표준화에 관한 연구. 한국방재학회논문집, 19(4), 143-149.
- 서민송, 유환희(2020), 공간계량분석 방법에 따른 시설물 화재 발생 유의성 분석. 한국측량학회지, 38, 281-293.
- 소방청(2021), 2020년도 화재통계연감
- 신봉희, 전해경(2018). 빅 데이터를 이용한 재해 정보 지원에 관한 연구. 한국융합학회논문지, 9(8), 25-32.
- 이동규(2016). 빅데이터 기반의 재난관리 시스템 운용 방향에 대한 예비적 고찰 - 미국, 영국, 한국의 사례 비교를 중심으로 -. 한국위기관리논집, 12(1), 17-32.
- 이상호(2011). 도시지역 특성과 화재발생의 상관성 연구 : 인구·사회·경제적 요인을 중심으로. 지방행정 정책연구, 1(1), 37-69.

- 이승운, 정승권(2021). 격자데이터 기반 잠재적 재해위험도 평가체계 구축. *Crisisonomy*, 17(1), 35-44.
- 이영진(2019). '김포 요양병원 화재' 2명 사망·47명 부상..합동감식, KBS News(2019. 09.24.)
- 이주호, 반용운, 김유미, 김원섭(2019). 도시공간 특성과 대형화재 발생의 인과지도 분석 : 도시의 사회, 경제, 환경적 요인과의 관계를 중심으로. *Crisisonomy*, 15(7), 1-15.
- 임현택, 배상영, 심규성, 김정수, 박성용, 김용성 (2016). 재해지도 활용성 증대를 위한 빅데이터 구축 및 적용 방안. *한국방재학회논문집*, 16(2), 205-213.
- 장경은, 정유미, 박초롱, 이종호, 손동욱(2021). GIS 네트워크 분석을 활용한 중구 화재 취약 지역 분석. *KIEAE Journal*, 21(6), 79-86.
- 전상수, 장현민(2008). 부산시 U-City 모델 구축을 위한 재해취약지 분석. *한국방재학회 논문집*, 8(2), 65-73.
- 정명균, 이석형, 김창수(2020). 빅데이터 분석을 활용한 재해 분야별 안전지수 서비스 모델 연구. *한국재난정보학회논문집*, 16(4), 682-690.
- 진승현, 이병훈, 김혜원, 구인혁, 권영진, 서동구(2021). 화재 확대 및 건축물 연면적을 고려한 건축물 용도별 화재위험분석에 관한 연구. *한국방재학회논문집*, 21(6), 141-148.
- 채수성, 장선연, 서동준(2018). 빅데이터 기반 재난 재해 위험도 분석 프레임워크 설계 및 구현. *디지털콘텐츠학회논문지*, 19(4), 771-777.
- 최명영(2018). 반복되는 대형화재의 시사점과 위험관리 개선방안, 화재보험협회
- 행정안전부(2018), 2017 재난안전통계연보
- 행정안전부(2021). 2020 재해연보. 행정안전부
- 황종아, 강지연, 김승주(2020). 서울시 도시화재 발생의 공간분석 및 도시쇠퇴 특성과의 관계. *주택도시연구*, 10(3), 1-20.
- 황성남, 조춘만(2011). GIS를 이용한 환경 위험과 인구적 특성의 관계에 대한 연구. *한국공간정보학회지*, 19(4), 1-10.
- André Boraks, Gregory M. Plunkett, Thomas Morris Doro, Frazer Alo(2021). Scale-Dependent Influences of Distance and Vegetation on the Composition of Aboveground and Belowground Tropical Fungal Communities. *Microbial Ecology* 81(80):1-10.
- Arik, S. O., & Pfister, T. (2020). TabNet: Attentive Interpretable Tabular Learning, Association for the Advancement of Artificial Intelligence.
- DreamQuark. (2022). PyTorch implementation of TabNet paper. GitHub.
- Madaio, Michael A(2018), "Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections", *Metro* 21.

□ 홈페이지

건설공사 안전관리 종합정보망, <https://www.csi.go.kr/hrp/stats/hrpStatsByAccdLis>

국가화재정보시스템 홈페이지, <https://nfds.go.kr/>

구글 홍수예측 시스템 관련 홈페이지,

https://support.google.com/publicalerts/topic/11030989?hl=en&ref_topic=2696427

<https://ai.googleblog.com/2020/09/the-technology-behind-our-recent.html>

<https://datastudio.google.com/reporting/9526ada3-68c8-4a5b-8661-4b1f61e6a6d5/page/CpieB?s=mhiHVW6FPES>

<https://www.theverge.com/2020/9/1/21410252/google-ai-flood-warnings-in-dia-bangladesh-coverage-prediction>

<https://egc.yale.edu/pande-and-coauthors-using-technology-save-lives-during-indias-monsoon-season>

네이버맵, <https://map.naver.com>

서부 펜실베이니아 지역 데이터 센터 홈페이지, <http://www.wprdc.org/>

서울 열린데이터 광장 홈페이지, <https://data.seoul.go.kr>.

안전정보 통합공개시스템 홈페이지, <https://safewatch.safemap.go.kr>.

안전정보 통합공개시스템 생활안전지도, <https://www.safemap.go.kr/main/smmap.do?flag=2>

텐서플로우 블로그, <https://tensorflow.blog>

Compmath 홈페이지, <https://compmath.korea.ac.kr/>

DEVOPS 홈페이지, <https://www.devops.ae/>

FireCast 홈페이지, <https://firecast.conservation.org>.

FireDanger Forecast 홈페이지,

<https://www.usgs.gov/fire-danger-forecast/about-fire-danger-forecast>

<https://firedanger.cr.usgs.gov/apps/gacc>

<https://firedanger.cr.usgs.gov/apps/staticmaps>

Github 홈페이지, <https://github.com/>

KOSIS 국가통계포털. <https://kosis.kr/>

Let the Machines Learn 홈페이지, <https://yashuseth.wordpress.com/>

Medium 홈페이지, <https://medium.com/>

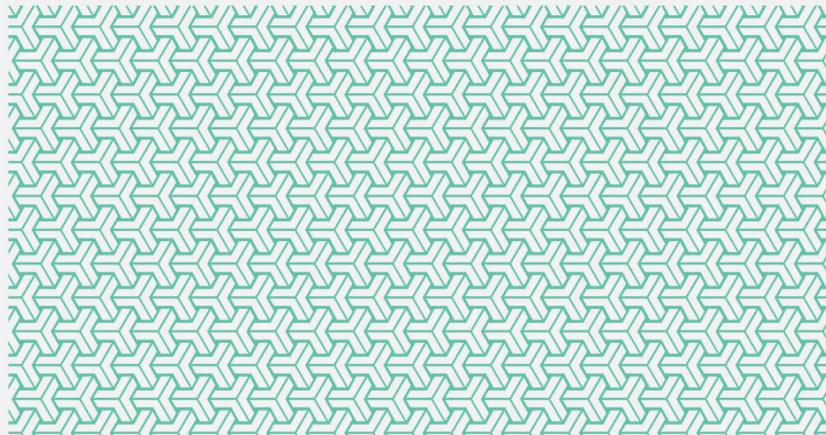
Metro21: Smart Cities Institute 홈페이지, <https://www.cmu.edu/metro21/index.html>

skikit-learn 홈페이지, <https://scikit-learn.org/stable/index.html>

WiliDocs 홈페이지, <https://wikidocs.net/>

A Study on the Development of Building Fire Prediction Model Based on Big-Data Analysis

Summary



Cho, Young Jin

Heo, Han Kyul

Ahn, Eui Soon

Ryu, Su Yeon

Song, Yu Mi

Hyeon, Tae Hwan

Safety accidents in buildings due to natural disasters such as earthquakes, floods, typhoons, and social disasters such as fires and crimes are increasing, and concerns about the large-scale human and property damage are also increasing. According to the "2020 Disaster Yearbook" published by the Ministry of Public Administration and Security, the amount of damage to buildings caused by typhoons, heavy rains and earthquakes between 2011 and 2020 was very high. In addition, there is a concern about large-scale human and property damage in the event of a safety accident due to an increase in medium and large-sized buildings. According to the Construction Safety Management Comprehensive Information Network, a total of 34,385 safety accidents occurred at construction sites in 2021, and a number of casualties from fires and collapses are continuously reported. The increase in aging buildings is accompanied by decrepit mechanical and electrical facilities of buildings, which is a factor that increases the risk of fire, so countermeasures are needed.

Accordingly, there is an increasing need for research on the use of data in architectural and urban spaces to solve disasters and disaster problems in buildings. Fire countermeasures for existing buildings are mainly to check fire prevention facilities to minimize damage in the event of a fire, and firefighters request the use of building drawing information in the electronic architectural administration information system to respond to fires. In light of the increasing amount of information on various and vast architectural and urban spaces such as the electronic architectural administration information system and the national fire information system, research is needed to establish a disaster and disaster prediction system based on big data analysis. With the development of fourth industrial revolution technologies such as AI, machine learning, and deep learning, the accuracy of predicting disasters and disasters is increasing significantly, and problem-solving skills using architectural and urban space data are secured, so research is needed to apply these skills.

Therefore, this study aims to establish a methodology for predicting dangerous areas to prevent safety accidents in buildings using attribute information of buildings and fire occurrence data in buildings, and to propose a model for predicting disaster and disaster risk areas. First of all, we intend to derive a linkage plan between fire occurrence data and electronic architectural administration information system data, and analyze it to establish a methodology for predicting fire occurrence risk areas. In particular, it is intended to derive a plan to link building fire occurrence data with electronic architectural administration information system data. In addition, we

would like to propose a disaster and disaster risk area prediction model. Based on the fire data risk area prediction methodology, we would like to propose a model that can predict disaster and disaster risk areas occurring in buildings and spaces such as collapse, heat wave, and flooding, and suggest a direction that can be used in policy.

Chapter 2 examines domestic and foreign disaster big data systems, big data analysis methodologies, and related research trends.

The state has established and operated various crisis management systems for disaster prevention, management, and recovery, and related systems are operated based on various disaster and disaster data. The disaster prevention and response ministries in Korea are overseen by the Disaster Safety Management Headquarters of the Ministry of Public Administration and Security, and operate the National Disaster Management Information System and the Integrated Safety Information Disclosure System to prevent and respond to disasters.

Fire and crime, which are social disasters, have established and operated individual systems. Fire is used as data for establishing and responding to fire prevention policies by providing information on the risk of fire occurrence by region through the National Fire Information System. The geographical profiling system operated by the National Police Agency mainly performs crime risk measurement for crime prevention. Overseas fire occurrence prediction systems primarily focus on fire occurrence and spread prediction systems through large-scale fire occurrence monitoring. Examples include FIRECAST and Fire Danger Forecast, which is operated by the U.S. Department of Interior.

The implications derived through the review of big data prediction systems related to disasters and disasters at home and abroad are that it is necessary to establish a system for collecting, processing, and analyzing related big data to prevent disasters and disasters. In addition, in order to respond quickly to disasters, take measures, and minimize damage, related data should be provided easily to policymakers and the general public, and building attribute information is important to establish a disaster and disaster prediction model for buildings.

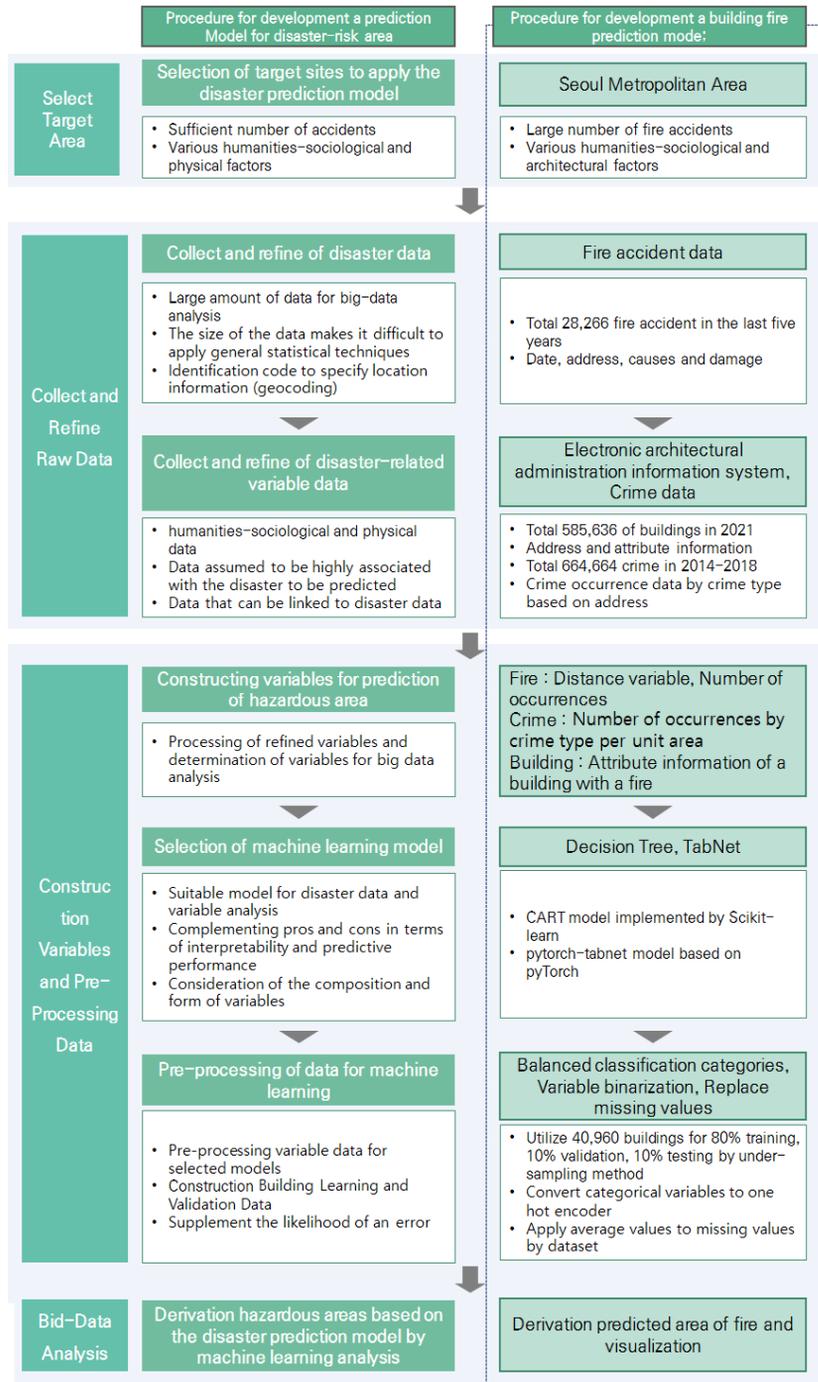
Looking at big data-based disaster and fire prediction research trends, the focus is on the types and lists of data that should be used for prediction and the big data research methodology that utilizes existing data. The type of data, the degree of construction, what data can be used other than disaster and disaster data, and the focus is on research on scenario construction for prediction.

In Chapter 3, related data selection and linkage measures were prepared to develop a building fire prediction methodology, and a building fire risk area prediction methodology was established.

A place with general characteristics of the city was selected so that it could be applied to various spatial units as a matter to consider when selecting a target site for the establishment of a prediction model for a fire hazard area. Therefore, in this study, Seoul was the target. Next, in order to apply to the big data analysis methodology, the fire data to be analyzed in this study was linked to the electronic architectural administration information system data. First of all, the structure of the fire occurrence data and the construction administration data was checked, and the PNU code containing address information was generated with the linkage KEY of the two data to perform the linkage.

After linking fire data and electronic architectural administration information system data, spatial informatization, or geocoding, was performed because distance variables will be introduced when analyzing big data. Building integration information was used as Basemap, and in the case of crime data, the data were enormous, so instead of extracting crime-prone buildings, it was derived in units of aggregated districts. As the next step, the ArcGis program was used to build a dataset for big data analysis, and three types of fire-related variables, six types of crime-related variables, and six types of building-related variables were built as independent variables, and whether a building fire occurred was built as dependent variables.

After setting variables for big data analysis, a machine learning model was selected. In this study, the CART model implemented by scikit-learn was used in consideration of the fact that fire was a binary variable. In addition, the TabNet model used in this study is a pyTorch-tabnet model based on pyTorch, and a classification model was used like the decision tree. After that, a total of 40,960 data were extracted as final analysis data by pre-processing and under-sampling according to each model, and 80% of the data to be analyzed was used as training data, 10% as verification data, and 10% as test data. In this way, the development of a prediction model for disaster and disaster risk areas was carried out in the order of data processing for analysis, variable construction, and big data analysis model selection to suit the model.



[Figure] Procedure for development prediction model for disaster-risk area

Source : Inhouse

In Chapter 4, the prediction model of buildings in dangerous areas was applied based on the construction of the final dataset and the selection of machine learning models performed in Chapter 3.

First of all, the current status of fire in buildings in Seoul was identified through technical statistics on building fire data. In the case of Seoul, the number of fires in apartments and detached houses was high, and in the case of total area, there were many fires in buildings exceeding 200m² and 500m² or less, and in the fire statistics by floor, there were many fires in low floors below five stories. In addition, as the aging progressed, the number of fires increased.

In big data analysis, the TabNet model was used to predict fire-prone buildings among decision tree models and deep learning models, and the types of fire-prone buildings and fire risk influencing factors were extracted through analysis of machine learning models. Through this process, the prediction model was verified and the characteristics of buildings vulnerable to fire were interpreted. As a result of the analysis through the decision tree model, among buildings with a total floor area of 5,160m² or more, the fire incidence rate in types excluding educational research facilities was the highest at 12.0%.

Through the extraction of fire risk influencing factors derived from the TabNet model, the possibility of using the predictive model through analysis of the deep learning model was verified. As a result of the analysis, the accuracy for the verification dataset was 67.5%, and the accuracy for the test dataset that was not used for learning and early interruption determination was 65.8%. As a result of extracting the factors influencing fire risk, the total area was the most relevant among the building properties, and the use was the next highest. It was found that the relationship between the violation building and the approval for use (oldness) was relatively low. In the case of crimes, the number of five major crimes was highly related, while the number of individual crimes was low. In the case of the distance variable, the radius of 1km of the fire area was higher than that of the surrounding 1km² and 500m.

The machine learning-based building disaster and disaster prediction model proposed in this study is a methodology that can identify fire vulnerability prediction, the type of vulnerability, and the cause of vulnerability judgment for each building unit. As the last step of the analysis, the fire vulnerability prediction result was visualized based on spatial information, and the applicability of GIS

visualization technology was reviewed.

Finally, the limitations of this study are that the analysis of all data was not performed due to some errors in fire occurrence data and electronic architectural administration information system data, and the prediction accuracy is relatively low because the types of variables applied in big data analysis are not diverse.

Therefore, in order to improve and upgrade the accuracy of the big data-based safety accident prediction model developed in this study, it is necessary to perform tasks to increase the integrity of related data and apply various humanities and sociological variables to increase the prediction accuracy.

Keywords :

Big-Data, Building Fire, Decision Tree, TabNet, Machine Learning