

건물별 화재 위험도 예측 및 분석: 재산 피해액과 화재 발생 여부를 바탕으로*

Risk Prediction and Analysis of Building Fires -Based on Property Damage and Occurrence of Fires-

이인아 · 오형록 · 이준기†

연세대학교 정보대학원

요 약

본 논문은 서울시에 존재하는 건물을 대상으로 화재 발생 시 재산 피해액, 화재 발생 여부를 예측하여 건물별 화재 위험도를 도출하였다. 본 연구는 건물의 특성뿐 아니라 해당 건물이 속한 행정동의 특성 및 소방 시설 접근성과 같은 변수를 활용하였다는 점에서 기존 선행연구와의 차이점을 지닌다. 앙상블 보팅(Ensemble Voting) 기법을 활용해 서로 다른 알고리즘을 병합했으며, 이를 통해 재산 피해액과 화재 발생 여부를 예측하고 변수 중요도를 추출하여 화재 위험도를 산출하는 방향으로 연구를 진행하였다. 구축된 모델을 사용하여 서울시에 존재하는 300개 건물을 대상으로 적용한 결과, 화재 위험도 1등급의 경우 건물 내 세대 수가 많으며, 관할 119안전센터가 가장 멀리 위치하는 등 화재 발생 시 그 규모를 키울 수 있는 요인들이 많은 것으로 나타났다. 반면 5등급의 경우, 주변 건물 수나 사업체 수는 많지만, 관할 119안전센터가 가장 가까이 위치해 있어 화재에 적절히 대응할 수 있는 건물들로 나타났다.

■ 중심어 : 화재 위험도, 재산 피해액, 화재 발생 여부, 기계학습, 앙상블 보팅(Ensemble Voting)

Abstract

This paper derives the fire risk of buildings in Seoul through the prediction of property damage and the occurrence of fires. This study differs from prior research in that it utilizes variables that include not only a building's characteristics but also its affiliated administrative area as well as the accessibility of nearby fire-fighting facilities. We use Ensemble Voting techniques to merge different machine learning algorithms to predict property damage and fire occurrence, and to extract feature importance to produce fire risk. Fire risk prediction was made on 300 buildings in Seoul utilizing the established model, and it has been derived that with buildings at Level 1 for fire risks, there were a high number of households occupying the building, and the buildings had many factors that could contribute to increasing the size of the fire, including the lack of nearby fire-fighting facilities as well as the far location of the 119 Safety Center. On the other hand, in the case of Level 5 buildings, the number of buildings and businesses is large, but the 119 Safety Center in charge are located closest to the building, which can properly respond to fire.

■ Keyword : Fire Risks, Property Damage, Occurrence of fires, Machine Learning, Random Forest

2021년 07월 23일 접수; 2021년 08월 13일 수정본 접수; 2021년 08월 19일 게재 확정.

* 본 연구는 2021년도 연세대학교 미래융합연구원(ICONS)의 지원에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

† 교신저자 (zlee@yonsei.ac.kr)

I. 서론

1.1 연구 배경 및 선행연구

2019년 화재 발생 건수와 화재 사망자 모두 전년도 보다 감소했으나 재산피해는 큰 폭으로 늘어났다. 소방청 화재통계연감에 따르면 연간 화재 발생(건)은 2017년(44,178), 2018년(42,338), 2019년(40,030)으로 감소하지만, 화재 재산피해(억원)는 2017년(5,069), 2018년(5,597), 2019년(8,059)으로 2019년도의 경우 전년도 대비 약 44.0%(2,462) 증가한 것으로 나타났다. 이는 대규모 재산피해를 불러온 화재가 여럿 발생했기 때문이며 초기 화재 진압의 중요성을 내포하고 있다.

2020년 5월, 화재 및 구급 활동 정보 데이터가 행정안전부 지원 국가중점 데이터 개방 분야로 선정되었다. 이는 국가적으로 활용 가치가 높은 화재 데이터를 적극적으로 개방하여 관련 분야 데이터 분석 산업의 활성화를 도모하려는 움직임으로 해석된다. 소방청은 화재 관련 데이터를 민간에 개방함으로써 관련 분야 연구개발에 크게 이바지할 수 있을 것이라는 기대감을 내비쳤다. 또한, 소방청 국가화재정보시스템에서는 전국 119상황실, 화재발생현황 데이터를 서울열린데이터광장에서는 원인별, 장소별 화재 발생통계, 비상 소화 장치 위치 정보 등의 데이터를 이미 제공하고 있어 화재 데이터에 대한 접근성이 나날이 높아지고 있으며 그 중요도 또한 인정받고 있는 바이다.

국내 중앙소방본부에서는 2017년 경기도 내 43만 개의 건물을 대상으로 화재 위험도를 예측하는 연구가 진행되었다. 해당 연구에서는 건물별로 화재 강도 및 화재 빈도를 각각 예측하여 최종적으로 화재 위험도를 도출한 후, 화재 위험도 등급별 건물 특성을 분석하였다. 화재 강도 예측에 사용된 독립 변수는 크게 건물의 건축요인, 화재요인, 피난 요인으로 나뉘며, 다중

선형회귀식을 통해 예측을 진행하였다. 화재 빈도 예측에 사용된 독립 변수는 크게 건물의 화재 원인, 예방 활동, 날씨 관련 요인으로 나뉘며, 의사결정트리 알고리즘 사용한 분류 예측 성능은 약 64%로 나타났다.

원종석(2017)은 화재 대응 취약성은 접근성, 대상물, 공간 특성과 연관 있다고 밝힌 바 있으며, 이는 초기 화재 진압과 관련된 변수로 판단된다. 이에 본 연구에서는 단순 건물 특성뿐 아니라 해당 건물이 속한 행정동 특성과 소방 시설 접근성 등의 변수를 새로 추가하여 건물별 화재 위험도를 새롭게 도출하고자 한다. 또한, 예측 결과에 대한 설명력 측면에서는 회귀 모델이 적합하지만, 선행연구에서 모델 자체의 성능이 좋지 않았기 때문에 화재 위험도에 관한 결과를 신뢰하기 어렵다는 한계점을 지녔다. 이러한 점을 보완하기 위해 본 연구에서는 의사결정 나무 기반 알고리즘을 사용하였으며 이를 앙상블 기법을 통해 모델을 구성하였다. 트리 기반 알고리즘은 예측 의사결정 과정에서 어떤 변수가 중요한 영향을 미치는지에 대한 상대적 변수 중요도를 도출할 수 있으므로 예측 성능과 모델 설명력 모두를 충족시키는 방법론이라 판단하였다.

1.2 연구 목적

본 연구는 서울시에 존재하는 건물을 대상으로 건물 특성과 건물이 속한 행정동 특성, 소방 시설 접근성 등의 변수를 종합적으로 반영하여 건물별로 화재로 인한 재산 피해액과 화재 발생 여부를 예측하고자 한다. 구축된 예측 모델을 통해 서울시 등록된 300개의 건물 데이터를 대상으로 재산 피해액과 화재 발생 여부 예측치를 각각 도출하고, 최종적으로 건물별 화재 위험도를 도출 및 등급화한다. 본 연구는 건물 특성뿐 아니라 소방 접근성 및 소속 지역 특성을 독립

변수로 반영한다는 점에 있어 기존 선행연구와 차이점을 보인다. 두 예측 모델 모두 상대적 변수 중요도를 도출하여 재산 피해액과 화재 발생 여부를 예측하는 데 있어 중요하게 작용하는 변수를 파악하고자 한다. 또한, 화재 위험도 등급별 건물 및 지역 특성 등을 분석하여 효과적인 화재 예방 활동을 위한 근거를 마련하는 목적을 지닌다.

II. 이론적 배경

2.1 화재 예측 관련 선행 연구

제공되는 공공 데이터를 바탕으로 화재 관련한 연구들이 활발히 진행되어오고 있으며 특히 건축물 정보(건축용도, 건축 자재 등)와 외부 환경적 요인(기상, 전기 등)을 활용한 화재 발생 예측 연구가 이루어지고 있다.

기계학습을 통한 전기화재 예측 모델 연구(고경석 외, 2018)에서는 한국전기안전공사, 기상청, 국토교통부, 소방본부 등의 기관에서 전기안전점검데이터, 전기화재사고정보 등의 데이터를 수집하였으며 랜덤 포레스트(Random Forest) 알고리즘을 통해 건물별 전기화재를 예측하였고 그 정확도는 약 74.7%를 보였다. 해당 연구에서는 절연저항 값, 습도, 풍속, 건축물 노후 연수, 용적률, 건폐율, 건축물 용도를 전기화재에 영향을 끼치는 주요 원인으로 선정하였다. 또한, 의사결정트리를 이용한 날씨에 따른 화재 발생 확률 예측 모델(류정우 외, 2014)에서는 국가화재정보시스템과 기상청에서 제공하는 데이터를 활용하여 날씨 조건을 봤을 때 화재 발생 확률을 예측하는 모델을 구축한 바 있다.

한편, 국내뿐 아니라 해외에서도 화재 관련 데이터를 활용한 연구 및 실 적용 사례가 존재한다. 미국 뉴욕시 소방청에서는 정부가 보유하고 있는 다양한 화재 관련 정보를 사용하여 건물별

화재 위험도를 예측하는 기계학습 모델을 개발하였으며, 각 자치구 내 소방서에서 가장 높은 위험도를 보이는 상위 15개 건물 목록을 제공하고 있다(이수진, 2019). 이는 인력 및 자원 부족으로 인해 매년 전체 건물 가운데 10%의 건물에 대해서만 점검이 이루어지고 있는 점을 보완하기 위한 접근으로, 위험도가 높은 건물을 선별하여 좀 더 효과적인 화재 방지 예방책을 세울 수 있다는 장점을 지닌다. 해당 모델에서는 건축물 화재 데이터, 세금, 보건위생 관련법 위반 내역, 건축물의 재질 등의 데이터를 활용하여 건물별 화재 위험도를 도출하였다.

지금까지의 선행연구를 살펴보면 주로 건물 건축양식, 화재 이력, 피난 설비, 건물 내 다중이용 업소 수 등과 같은 건축물 자체 특성과 날씨, 전기 요인 등의 외부 환경적 요인을 예측 모델에 사용하였으며, 해당 건물이 속한 지역 특성 등과 같은 거시적 접근이 부족하다는 한계점을 지닌다.

특히, 국내 중앙소방본부에서 건물별 화재 강도를 예측하는 데 있어 건물 단위 정보만 활용하기에는 주변 환경으로 인한 화재 규모의 변동성을 충분히 설명하기 어렵다. 실제로 종속 변수와 독립 변수 사이의 상관관계는 매우 낮았으며 다중선형회귀식을 사용한 결과 그 예측 성능이 좋지 않았다고 밝혔다.

2.2 랜덤 포레스트(Random Forest)

랜덤 포레스트(Random Forest) 알고리즘은 주어진 데이터의 리샘플링(Resampling)을 통해 생성한 다수의 의사결정나무 알고리즘을 결합하여 정확도를 높이는 기법이다. 의사결정나무 알고리즘은 데이터 사이의 관계와 규칙을 파악하고 모형화하는 분석기법으로 정규성, 등분산성, 선형성 가정이 필요 없는 방법론이다(박명화 외, 2013). 랜덤 포레스트는 이러한 의사결정나

무를 여러 개 파생시켜 그 결과를 종합하여 제공하며 의사결정나무의 과적합 문제를 개선한 알고리즘이다. 랜덤 포레스트는 각 의사결정나무를 만들 때 무작위로 학습 데이터셋과 독립 변수를 선정하여 예측을 진행하며, 개별적인 정확도는 떨어질 수 있으나 모든 의사결정나무를 종합하여 예측을 수행하므로 정확도와 안정성이 높아진다는 장점을 지닌다(김성진, 안현철, 2016).

2.3 XGBoost

부스팅(Boosting) 기법은 초기 샘플 데이터를 사용하여 약한 학습모델을 생성하고 학습 결과의 오차를 줄이는 방향으로 신규 학습모델을 반복적으로 추가하는 앙상블 기법이다(오재영, 함도현, 이용건, 김기백, 2019). 특히 기울기 부스팅(Gradient Boosting)은 이전 학습모델의 잔차를 예측하는 신규 모델을 계속 추가해 나아가는 알고리즘인데(Ogunleye, Wang, 2019), 학습 속도가 느리며 과대 적합(Overfitting)의 문제를 발생시키는 단점이 있다. 이를 보완하기 위해 등장한 알고리즘이 XGBoost이다. XGBoost는 의사결정나무 기반 알고리즘이며 기울기 부스팅 구조를 사용한다. XGBoost는 병렬 처리, 규제 등을 통해 훈련손실을 최소화하면서 과적합을 줄이는 최적화된 모델을 생성한다(황혜진, 김수현, 송규원, 2018).

2.4 LightGBM

LightGBM은 2016년 마이크로소프트 MSRA에서 설계한 효율적인 GBDT(Gradient Boosting Decision Tree) 기반 알고리즘이다(Ma, Sha, Wang, Yu, Yang, Nju, 2018). 기존 GBDT 기반 알고리즘(e.g. XGBoost, pGBRT 등)은 모든 가능한 분할지점에 대한 정보 이득(Information gain) 평가를 위해 데이터 전부를 확인하기 때문

에 대용량 고차원 데이터에서는 성능을 제대로 발휘하지 못한다. 정보 이득이란, 특정 속성을 선택함으로써 데이터를 더 잘 구분 짓는 것을 말한다. LightGBM은 GOSS(Gradient-based One-Side Sampling)와 EFB(Exclusive Feature Bundling) 기법으로 문제를 해결하였다.

GBDT 알고리즘에서 기울기(Gradient)가 큰 데이터 속성이 정보 이득 관점에서 더욱 큰 역할을 한다. GOSS는 기울기(Gradient)가 큰 데이터 속성은 유지하고 기울기(Gradient)가 작은 데이터 속성은 일정 확률로 무작위로 제거하는 기술이다. EFB는 변수 개수를 줄이기 위해 희소한 변수 공간의 특성에 따라 상호배타적인 변수들을 묶는 기법이다(Ke, Meng, Finley, Wang, Chan, Ma Ye, Liu, 2017).

2.5 Ensemble Voting

앙상블은 서로 다른 분류기를 하나의 메타 분류기로 합쳐 개별 분류기로 독립된 성능을 발휘할 때보다 더 나은 일반적 성능을 낼 수 있도록 하는 기법이다. 앙상블 보팅(Ensemble Voting)은 서로 다른 단일 분석기법들이 도출한 결과에서 투표를 통해 최적의 결과를 선정하는 앙상블 기법인데, 크게 하드 보팅(Hard Voting)과 소프트 보팅(Soft Voting)으로 구분할 수 있다(김한민, 2020). 하드 보팅(Hard Voting)은 개별 분류기의 예측 결과에서 다수결을 통해 최종 예측이 이루어지지만, 소프트 보팅(Soft Voting)은 각 분류기가 제시한 예측 결과 확률의 평균값을 바탕으로 최종 결과를 선정한다. 그리고 본 연구에서는 소프트 보팅(Soft Voting)을 이용하여 각각의 모델을 앙상블 하였다.

III. 연구 방법

3.1 변수 선정 및 데이터 수집

재산 피해액과 화재 발생 여부 각각 나누어서 해당 종속 변수와 관련 있는 변수를 탐색 및 데이터를 수집하였다. 데이터는 모두 서울열린데이터광장에서 수집했으며, 본 분석에서는 2018년 데이터를 사용하였다. 독립 변수는 크게 건물 특성, 건물 소속 행정동 특성, 소방 시설로 나뉘며 종속 변수는 2018년 서울시에서 실제 화재가 발생한 건물의 재산 피해액을 사용하였다.

본 연구는 재산 피해 규모를 예측하기 위해 선행연구에서 사용된 변수 이외의 영향 요인에 대한 추가적인 탐색을 진행하였으며 크게 ① 소방차 진입에 어려움을 유발하는 변수와 ② 물 관할 소방서 정보 및 소방 시설 접근성 변수로 나뉜다.

3.1.1 재산 피해액

소방차 진입에 어려움을 유발하는 변수로는 건물이 속한 행정동 내 도로 폭이 4m/3m/2m 미만인 구간의 비율과 길이 총합 변수, 행정동 내 야외 공용주차장 수 및 주차 가능 수를 추가하였다. 화재 발생 시 골든 타임은 화재 발생 후 5분 이내이지만, 주변 주차장 문제 및 좁은 도로 폭으로 인한 진입의 어려움이 도착 시간을 지연시켜 대형화재로 번지는 사례가 다수 존재한다(황의홍 외, 2018). 주택건설기준에서는 소방차 최소 진입 도로 폭을 4m로 지정하였으며, 소방청에서는 도로 폭 3m는 소방차 진입 곤란 지역, 2m는 진입 불가 지역으로 분류한 기준에 따라 도로 폭 4m/3m/2m 미만인 구간의 비율과 해당 구간의 길이 총합을 변수로 추가하였다. 또한, 야외에 주차된 차가 많은 경우도 소방차 진입을 방해할 수 있으므로 행정동 내 공용주차장 수와 주차 가능한 차량 수 또한 변수로 추가하였다.

건물 관할 소방서 정보 및 소방 시설 접근성

변수로는 해당 건물을 담당하는 소방서 및 119 안전센터와의 거리와 해당 안전센터의 담당하고 있는 관할 면적, 관할 인구 변수를 추가하였다. 각 건물을 담당하는 소방서와 119안전센터의 관할 면적이 넓거나 관할 인구 수가 많을수록 화재 발생 시 진압을 위한 자원이 부족하여 그 규모가 커질 수 있기 때문이다. 또한, 해당 건물이 속한 행정동 내 소방용수시설과 비상소화장치의 개수를 소방 시설 접근성을 나타내는 변수로써 사용하였다.

최종적으로 종속 변수인 재산 피해액을 기준으로 이상치를 제거한 후 총 423개의 건물 데이터를 예측 모델에 사용하였다.

3.1.2 화재 발생 여부

건물의 화재 발생 자체에 영향을 줄 수 있는 독립 변수는 크게 건물 특성, 에너지 사용량, 건물 소속 행정동 특성으로 나눌 수 있다. 종속 변수는 2018년 서울에서 실제 화재가 발생한 건물 423건과 화재가 발생하지 않은 건물 423건을 무작위 추출하여 클래스별 데이터 수를 통일한 846개의 건물 데이터를 사용하였으며 화재 발생 건물은 1, 화재 미발생 건물은 0으로 설정하였다.

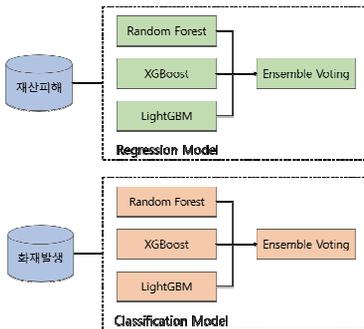
본 연구에서는 기존 선행연구에서 사용한 변수 이외에도 실제 유동 인구수를 반영하기 위해 생활인구 변수를 추가하였으며, 건물 용도 및 주변 사업체 수 또한 화재 빈도에 영향을 줄 수 있으므로 사업체 수 및 건물 수 변수를 반영하였다.

화재 발생 여부에 사용된 변수 가운데 건물 연식, 건물 구조 재료, 건축 면적, 건물용도, 건물 높이 변수는 재산 피해액 예측에도 사용된 변수로, 피해 금액과 화재 발생 여부를 예측하는 데 공통으로 사용되었다. 해당 변수들은 모두 건물 특성을 나타내는 변수로, 화재를 유발할 수 있는 요인임과 동시에 화재 규모를 키울

수 있는 변수로도 작용할 수 있으므로 공통 변수로 반영하였다.

3.2 모델링

본 연구에서 재산 피해액은 회귀 알고리즘을, 화재 발생 여부는 이진 분류 알고리즘으로 예측 모델을 구성하였고 다음 그림 1과 같다. 총 3종류의 알고리즘(Random Forest, XGBoost, LightGBM)을 선택했으며 이를 앙상블 하여 최종 모델을 구축하고 예측을 진행 하였다. 그리고 이는 서로 다른 알고리즘의 결과를 반영함으로써 출력 결과의 객관성을 확보하기 위한 시도로 예측값이 편향되지 않도록 하기 위함이다.



〈그림 1〉 Prediction Modeling

3.3 분석 방법

본 연구에서 화재 위험도는 중앙소방본부 (2017)에서 사용한 화재 위험도를 참고 및 활용 하였다.

재산 피해액 X_α 는 예측 모델을 통해 도출된 건물별 재산 피해액 예측 금액에 범위 표준화를 진행한 후 100을 곱하여 0~100 사이의 지수로 산출하였다.

$$X_\alpha = standardization(\hat{y}_{재산피해금액}) \times 100 \quad (1)$$

화재 발생 여부 예측 모델은 건물별 화재 발생 확률을 0~1 사이의 값으로 제공하므로, 화재 발생 여부 지수(X_β)는 예측확률에 100을 곱하여 0~100사이의 지수로 산출하였다.

$$X_\beta = \hat{y}_{화재발생확률} \times 100 \quad (2)$$

마지막으로 재산 피해를 지수와 화재 발생 여부 지수를 곱한 값에 제곱근을 취해 최종적으로 화재 위험도를 도출하였으며, 위험도 점수대별 등급표는 <표 1>과 같다.

$$Y_{화재위험도} = \sqrt{X_\alpha \times X_\beta} \quad (3)$$

(0 ≤ Y_{화재위험도} ≤ 100)

〈표 1〉 화재 위험도 등급표

등급	위험도 점수
1	70점 이상
2	55점 이상 70점 미만
3	40점 이상 55점 미만
4	25점 이상 40점 미만
5	25점 미만

IV. 분석 결과

4.1 화재 발생 여부

화재 발생 여부는 화재가 발생한 건물은 1, 화재가 발생하지 않은 건물은 0으로 분류하는 예측 문제이며, 본 연구에서 모델의 분류 성능 지표로 Accuracy를 사용하였다. 기존 연구와의 예측 성능 비교는 <표 2>와 같으며, 랜덤 포레스트의 74.7%보다 약 3% 더 높은 정확도를 확인할 수 있었다.

화재 발생 여부 예측 모델의 변수 중요도를 도출한 결과는 다음 <표 3>과 같으며, 건축 면적 변수가 전체 변수 가운데 약 10%로 가장 중요한

변수로 선정되었다.

<표 2> 모델 예측 성능 비교

단위: %

독립변수	Method	Acc.
건물특성	Decision Tree	64.2
건물특성, 기상요인, 전기요인	Random Forest	74.7
건물특성, 전기요인, 행정동 특성, 인구특성	Ensemble Voting (RF, XGB, LGBM)	77.3

<표 3> 화재 발생 여부 변수 중요도

변수	중요도
건축면적(m ²)	0.097
건물 용도	0.081
연면적(m ²)	0.061
세대수	0.04
건물 연식	0.03

4.2 재산 피해액

재산 피해액 모델의 예측 성능으로는 RMSE를 사용하였다. RMSE는 실제값(y_i)과 모델의 예측값(\hat{y}_i)의 차이를 다룰 때 흔히 사용되는 척도이다. 알고리즘별 모델 성능 결과는 <표 4>와 같으며 Ensemble Voting 모델 예측 결과 RMSE 값은 약 265로 도출되었다. 이는 평균적으로 재산 피해 금액의 실제값과 예측값이 약 26만 5천 원 정도의 편차가 존재함을 의미한다.

<표 4> 회귀 모델 성능 비교

단위: 천원

Algorithm	Method	RMSE
Bagging	Random Forest	271.4
Boosting	XGBoost	331.1
Boosting	LightGBM	284
Ensemble	Ensemble Voting (RF, XGB, LGBM)	265.5

재산 피해 금액 예측 모델의 변수 중요도를 도출한 결과는 다음 <표 5>와 같으며, 일반음식점 수 변수가 전체 변수 가운데 약 25%로 가장 중요한 변수로 선정되었다.

<표 5> 재산 피해액 변수 중요도

변수	중요도
일반음식점 수	0.249
안전센터거리	0.068
건축면적(m ²)	0.051
행정동 도로폭 3m미만 구간 길이	0.045
위험물저장 및 처리시설	0.033

4.3 화재 위험도 도출

모델 학습에 사용되지 않은 300개의 건물 데이터를 대상으로 재산 피해액, 화재 발생 여부를 각각 예측하였으며 최종적으로 화재 위험도 점수를 도출 및 등급을 매긴 결과는 <표 6>과 같다. 해당 결과는 5등급으로 갈수록 화재 위험도가 낮은 건물을 의미한다. 화재 위험도 1등급에 해당하는 건물의 수는 25개이며, 해당 건물들의 평균 화재 발생 확률은 약 80%, 화재 발생 시 재산 피해액 예측치는 평균 49만원으로 도출되었다. 반면 화재 위험도 5등급에 해당하는 건물의 수는 50개이며 해당 건물들의 평균 화재 발생 확률은 약 37%, 화재 발생 시 재산 피해액 예측치의 평균은 약 10만원으로 도출되었다.

<표 6> 화재 위험도 등급 도출 결과

화재 위험도 등급	건물 수	재산 피해액 평균 예측치(천원)	화재 발생 확률 평균 예측치(%)
1	25	490.39	85.32
2	53	284.06	79.6
3	87	195.83	70
4	85	153.98	61
5	50	94.4	37.3

<표 7>는 화재 위험도 등급별 건물 특성을 나타낸 것으로 해당 건물들이 지닌 수치들의 단순 평균치이다. 화재 위험도 1등급 건물의 경우 대지 면적/연면적/세대 수/높이/지상층 수의 평균이 가장 높았으나 건물연식은 오히려 가장 낮았다. 세대 수, 건물 높이 모두 위험도가 높아질수록 수치가 커졌으며 건물 연식은 위험도가 높아질수록 수치가 작아지는 양상을 보였다. 즉, 화재 위험도 5등급에 속한 건물들은 건물 내 세대수가 가장 적게 나타나는 것으로 보아(5.08) 사

람이 적을수록 건물 내 화재가 발생할 확률 자체가 낮아지는 것으로 해석할 수 있다. 반면 화재 위험도가 높은 건물은 비교적 최근에 지어졌으며 많은 사람이 거주할 수 있는 대규모 건물로 거주 인구가 많음에 따라 화재가 일어날 가능성 또한 높은 것으로 해석된다.

<표 8>은 화재 위험도 등급별 소속 행정동 특성을 나타낸 것으로 해당 건물들이 속한 행정동 지표들의 단순 평균치이다. 화재 위험도 1등급 건물들이 속한 행정동의 인구 관련 특성들이 대

<표 7> 화재 위험도 등급별 건물 특성

화재 위험도 등급	건축면적 (m ²)	대지면적	연면적	세대 수 (세대)	높이 (m)	지상 층수	지하 층수	비상용 승강기 수	건물 연식
1	448.19	2455.23	9388.98	64.84	24.63	9.76	1,6	0.2	18.12
2	568.97	964.63	8034.05	63.15	19.03	8.26	1.02	0.26	20.68
3	461.32	691.87	5843.86	56.32	23.32	9.12	1.03	0.24	21.30
4	275.15	401.32	2065.37	20.78	12.61	5.54	0.95	0.17	22.02
5	108.45	102.03	621.34	5.08	8.31	3.6	0.72	0.17	26.42

<표 8> 화재 위험도 등급별 행정동 특성

화재 위험도 등급	도로 폭 4m 미만 구간 길이 합계(m)	도로 폭 3m 미만 구간 길이 합계(m)	도로 폭 2m 미만 구간 길이 합계(m)	도로 폭 4m 미만 구간 길이 비율(%)	도로 폭 3m 미만 구간 길이 비율(%)	도로 폭 2m 미만 구간 길이 비율(%)	단독주택	건물 수	공영 주차장 수
1	613.68	537.66	501.58	0.24	0.12	0.03	3691	6554	6.6
2	623.58	644.69	516.38	0.25	0.12	0.05	3317	6018	7.6
3	591.96	568.81	540.52	0.35	0.16	0.04	2780	5147	7.7
4	610.51	555.50	956.59	0.36	0.15	0.03	2945	5579	9.1
5	829.21	826.05	1519.91	0.43	0.19	0.06	2658	4896	9.0

화재 위험도 등급	시간대별 생활인구 평균(명)	인구(명)	면적대비 생활인구 밀도	연간 가스 사용량	연간 전기 사용량	전체평균 연령	65세이상 인구 수	65세이상 인구 비율	행정동 면적(km ²)
1	61,371,729	27,706	7,662,286	341,313,603	282,419,255	41.72	3724.68	0.14	2.6
2	56,725,709	24,719	8,490,096	307,696,563	241,786,558	42.64	3554.36	0.14	1.6
3	58,418,897	26,361	8,839,255	235,074,876	233,651,262	42.53	3727.77	0.15	1.4
4	64,593,102	27,287	10,029,790	217,404,086	251,482,329	42.37	3724.51	0.15	1.3
5	70,138,320	27,926	10,794,759	215,960,481	298,196,694	42.46	3711.37	0.15	1.3

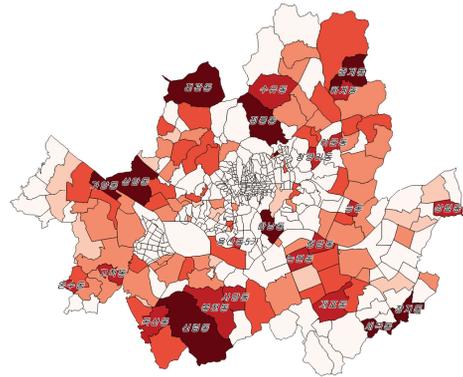
부분 낮은 수치를 보이지만 연간 가스 사용량은 상대적으로 높다는 것을 확인할 수 있다. 즉, 화재 위험도가 높을수록 1인당 전기/가스 사용량이 많으며 이는 곧 화재로 이어질 수 있는 위험을 내포하고 있다. 또한, 화재 위험도 1등급 건물들이 속한 행정동들의 평균 면적이 가장 넓었고 5등급으로 갈수록 좁아지는 양상을 보였다.

한편, 초기 화재 진압의 중요성을 앞서 언급한 바 있으며 이와 관련한 변수를 살펴보면 1등급을 받은 건물의 경우, 도로 폭 길이 합계와 비율이 낮았으며 면적 대비 도로의 비율 또한 가장 낮은 것으로 확인된다. 등급별 건물 주변 소방 시설 접근성(<표 9>)에서 5등급은 소방서, 119안전센터와의 거리가 다섯 등급 가운데 가장 가까웠으며, 해당 센터가 담당하는 면적 또한 가장 좁은 것으로 나타나 건물에 화재가 발생하더라도 이를 해결할 수 있는 소방 자원이 충분히 배치된 것으로 해석할 수 있다. 반면, 1등급을 받은 건물은 전반적으로 관할 소방서, 119안전센터와의 거리가 멀게 위치해있으며 관할 면적도 가장 넓게 나타나 해당 건물을 관할하는 센터들의 소방 자원이 다른 등급의 건물에 대비해 부족함을 보여준다. 그리고 이는, 초기 화재 진압을 위한 주변 소방 시설 접근성이 취약하여 화재 발생 시 그 강도를 키울 수 있는 요인으로 작용할 수 있다.

최종적으로 서울시 300개 행정동 대상으로 화재 위험도 등급을 평균하여 집계한 결과는 다

음 <그림 2>과 같다.

화재 위험도 1등급의 경우, 전광동, 정릉동, 중계동으로 2등급은 사당동, 논현동, 개포동 등이 있으며 서울시 300개 행정동에 대한 결과는 <표 10>에 정리하였다.



<그림 2> 서울시 행정동별 화재 위험도 등급

V. 결 론

분석 결과, 재산 피해액, 화재 발생 여부 예측에 있어 건물의 건축 면적 변수가 모두 중요한 변수로 도출 되었으나, 화재 발생 여부에서는 건물의 특성이, 재산피해액에서는 주변 소방 시설에 대한 접근성 정도가 중요하게 도출되었다는 차이점을 보였다. 또한, 화재 위험도 1등급에 속하는 건물들은 전반적으로 세대 수가 많아 화재가 일어날 가능성이 크면서도 행정동 내에 화

<표 9> 화재위험도 등급별 주변 소방시설 접근성

화재 위험도 등급	소방서 거리	119 안전센터 거리	119 안전센터 관할 인구 수	119 안전센터 관할면적	비상소화장치 수	소방용수시설 수
1	3.776	2.172	123108.56	9.7812	32.9106667	630.123
2	3.8	1.86981132	112244	6.05056604	20.9992662	578.413365
3	3.27471264	1.81149425	103493.034	5.61931034	11.0470215	499.517864
4	3.32588235	1.53882353	102860.424	4.39647059	13.8389122	553.235212
5	3.444	1.35	90402.52	4.6002	16.7755556	577.0575

〈표 10〉 화재 위험도 등급별 서울시 행정동

화재 위험도 등급	서울시 행정동
1	장지동, 세곡동, 가양동, 상암동, 진관동, 정릉동, 동소문동4가, 한남동, 신림동, 중계동
2	논현동, 봉천동, 상일동, 청담동, 온수동, 고척동, 하계동, 수유동, 이문동, 청량리동, 능동, 용산동5가, 독산동, 사당동, 개포동
3	일원동, 삼성동, 가산동, 시흥동, 방배대오, 등촌동, 상계동, 월계동, 도곡동, 대치동, 대방동, 궁동, 신도림동, 마곡동, 성산동, 망원동, 홍은동, 창천동, 대신동, 홍제동, 응암동, 구산동, 불광동, 상월곡동, 하월곡동, 보문동6가, 동소문동1가, 신내동, 목동, 장안동, 화양동, 하왕십리동, 이촌동, 문배동, 원효로1가
4	성내동, 삼전동, 서초동, 신길동, 염창동, 공릉동, 쌍문동, 미아동, 성수동2가, 부암동, 면목동, 노량진동, 화곡동, 개봉동, 자양동, 마천동, 잠실동, 도림동, 영등포동, 오류동, 내발산동, 목동, 합정동, 동교동, 서교동, 아현동, 북가좌동, 녹번동, 방학동, 번동, 장위동, 보문동2가, 망우동, 중화동, 구의동, 금호동1가, 동자동, 신정동, 구로동, 신사동, 상도동, 중곡동
5	고덕동, 가락동, 본동, 방화동, 신월동, 역삼동, 천호동, 길동, 가리봉동, 공향동, 노고산동, 신공덕동, 공덕동, 역촌동, 수색동, 동소문동6가, 신계동, 신당동

재 관련 인프라가 충분하지 않고 진입에 어려운 환경으로 인해 화재 발생 시 그 피해 규모가 커질 수 있는 위험을 지닌 건물들로 설명할 수 있다. 반면 화재 위험도 5등급에 속하는 건물들은 소속 행정동 내 사업체 수나 생활인구 등이 많아 화재 발생 가능성이 확연히 낮은 편이 아니지만, 소방서 및 안전센터가 비교적 인근에 위치해 화재 발생 시 신속하게 대응이 가능한 건물들이다. 실제로 화재 위험도 5등급 건물들의 화재 발생 예측 평균치는 약 37.3%로 절반보다 낮은 수치이며, 재산 피해액 예측 평균치는 약 10만원으로 가장 낮다는 점을 통해 확인할 수 있다.

화재라는 것은 대부분이 인체에 의한 것이기 때문에 화재 발생 자체를 완전히 막을 수는 없지만, 발생하더라도 그 피해 규모를 최대한 줄일 수 있도록 주변 인프라 및 소방력을 강화할 필요성이 존재한다. 그 때문에 화재 위험도가 높게 책정된 건물의 경우, 지속적인 예방 활동과 함께 주변 소방 관련 시설을 늘리고 관할 소방서/안전센터를 조정하거나 그 자원을 효율적으로 배분할 수 있도록 해야 한다. 또한, 화재 발

생 여부에 있어 건물 내 세대 수가 중요하게 도출된 만큼, 정기적인 화재 예방 교육 등과 같은 화재 발생을 미연에 방지할 수 있는 노력 또한 필요하다.

본 연구에서는 건물별로 화재 위험도를 예측하는 데 있어 건물 특성뿐 아니라 해당 건물이 속한 행정동의 소방 시설 및 접근성 등을 새로운 변수로 고려하여 그 중요성을 확인하였다는 점에서 시사점을 지닌다. 특히, 화재 규모에 있어 좁은 도로 폭, 주변 소방 시설 개수, 관할 소방서/안전센터 등이 중요한 변수로 나타났으므로, 해당 변수를 기반으로 그 규모를 줄이기 위한 예방 활동을 사전에 진행해야 한다. 그러나 화재로 인한 재산 피해액 예측에 있어, 실제 피해 규모 금액의 편차가 컸기 때문에 예측 정확도가 화재 발생 여부 예측 결과와 비교하여 다소 떨어졌다. 이 때문에 향후 연구에서는 피해 규모에 영향을 주는 요인을 더 넓은 관점에서 모색하여 데이터를 수집하고, 피해 예측액을 정확히 맞추는 데 초점을 두기보다는 그 규모가 크고 작을지를 대략적으로 판단하여 예측 결과를 활용하는 것이 중요하다고 판단된다.

참 고 문 헌

[1] 권수현, “작년 화재 발생·사망자 감소…재산피해는 역대 최대”, 『연합뉴스』, 2020.01.06. 행정안전부, “행안부, 디지털 뉴딜 선도할 25개분야 공공데이터 개방”, 『대한민국 정책브리핑』, 2020.08.18.

[2] 원종석, “서울시 화재 대응력 향상방안”, 『서울연구원 정책리포트』, 제239호(2017), pp.1-21.

[3] 고경석, 황동현, 박상준, 문가경, “기계학습을 통한 전기화재 예측 모델 연구”, 『한국정보전자통신기술학회지』, 제11권, 제6호(2018), pp.703-710.

[4] 류정우, 김영진, 김은주, 김명원, “기상예보기반 화재발생 확률 예측모델의 생성 기법”, 『정보과학회논문지』, 제20권, 제2호(2014), pp.68-78.

[5] 이수진, “화재 위험 큰 건물 찾아내는 머신러닝 예측모델 구축”, 『서울연구원 세계도시동향』, 제452호(2019).

[6] 박명화, 최소라, 신아미, 구철화, “기상예보기반 화재발생 확률 예측모델의 생성 기법”, 『Journal of Korean Academy of Nursing』, 제43권, 제1호(2013), pp.1-10.

[7] 김성진, 안현철, “기업신용등급 예측을 위한 랜덤 포레스트의 응용”, 『산업혁신연구』, 제32권, 제1호(2016), pp.187-211.

[8] 류정우, 김영진, 김은주, 김명원, “기상예보기반 화재발생 확률 예측모델의 생성 기법”, 『정보과학회논문지』, 제20권, 제2호(2014), pp.68-78.

[9] 오재영, 함도현, 이용건, 김기백, “XGBoost 기법을 이용한 단기 전력 수요 예측 및 하이퍼파라미터 변화에 따른 영향 분석”, 『전기학회논문지』, 제68권, 제9호(2019), pp.1073-1078.

[10] Qgunleye, A., and Wang, Q. G., “XGBoost Model for Chronic Kidney Disease Diagnosis”, 『IEEE/ACM transactions on computational biology and bioinformatics』, Vol.17, No.6(2019), pp.2131-

2140.

[11] 황혜진, 김수현, 송규원 “XGBoost 모델 해석을 통한 노인의 인지능력 개선·악화 요인 탐구”, 『차세대컴퓨팅학회지』, 제14권, 제3호(2018), pp.16-24.

[12] Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., and Niu, X., “Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGBoost algorithms according to different high dimensional data cleaning”, 『Electronic Commerce Research and Applications』, Vol.31(2018), pp.24-39.

[13] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ..., and Liu, T. Y., “Lightgbm: A highly efficient gradient boosting decision tree”, 『Advances in neural information processing systems』, Vol.30(2017), pp.3146-3154.

[14] 김한민, “양상블 머신러닝 기법과 블록체인 정보를 활용한 이더리움 영클 블록 예측 분석”, 『디지털융복합연구』, 제18권, 제11호(2020), pp.129-136.

[15] 황의홍, 최지훈, 최돈묵 “소방차 출동 시 효율적인 골든타임 확보 방안에 관한 연구”, 『한국방재학회지』, 제18권(2018), pp.119-126.

저 자 소 개



이 인 아(Ina Lee)

- 2018년 : 국민대학교 경영학과 (경영학사)
- 2020년 : 연세대학교 정보대학원 (정보시스템학 석사)
- 관심분야 : 빅데이터 분석, 모델링, NLP



오 형 록(Hyung-Rok Oh)

- 2020년 : 아주대학교 E-business학과 (경영학사)
- 2020년 ~ 현재 : 연세대학교 정보대학원 비즈니스 빅데이터 분석 트랙 (석사과정)
- 관심분야 : 빅데이터 분석, 데이터 마이닝, 머신러닝 등

데이터 마이닝, 머신러닝 등



이 준 기(Zoonky Lee)

- 1985년 : 서울대학교 컴퓨터사이언스(학사)
- 1991년 : 카네기멜론대학 사회심리학(석사)
- 1999년 : 남가주 대학교 경영정보학(박사)
- 2004년~현재 : 연세대학교 정보대학원 교수
- 관심분야 : Big Data Analytics, Digital Transformation, Open Collaboration 등