

## 기계학습을 통한 전기화재 예측모델 연구

고 경 석\*, 황 동 현, 박 상 준, 문 가 경

### Electrical fire prediction model study using machine learning

Kyeong-Seok Ko\*, Dong-Hyun Hwang, Sang-June Park, Ga-Gyeong Moon

**요 약** 매년 전기화재사고에 대한 사고유형 분석, 점검 등 전기적 화재사고를 줄이기 위해 다양한 노력이 있었으나, 효율적인 의사결정지원 체계 및 기존 누적 데이터 활용방안의 미비로 효과적인 대처방안이 부재한 현황이다. 본 연구는 전기안전점검데이터, 전기화재사고정보, 건축물정보, 기상청정보 등 데이터 기반의 전기화재를 예측하는 알고리즘을 개발하고 이를 활용하여 전기화재사고를 줄이는데 목적이 있다. 본 연구에서는 한국전기안전공사, 기상청, 국토교통부, 소방본부 등 기관별로 수집된 데이터를 전처리, 융합, 분석, 모델링, 검증 과정을 거쳐 전기화재에 영향을 끼치는 요인과 예측모델을 도출하였다. 주요요인으로 절연저항 값, 습도, 풍속, 건축물 노후년수, 용적율, 건폐율, 건축물용도로 나타났고, Random forest 알고리즘을 활용한 예측모델은 74.7%의 정확도를 얻었다.

**Abstract** Although various efforts have been made every year to reduce electric fire accidents such as accident analysis and inspection for electric fire accidents, there is no effective countermeasure due to lack of effective decision support system and existing cumulative data utilization method. The purpose of this study is to develop an algorithm for predicting electric fire based on data such as electric safety inspection data, electric fire accident information, building information, and weather information. Through the pre-processing of collected data for each institution such as Korea Electrical Safety Corporation, Meteorological Administration, Ministry of Land, Infrastructure, and Transport, Fire Defense Headquarters, convergence, analysis, modeling, and verification process, we derive the factors influencing electric fire and develop prediction models. The results showed insulation resistance value, humidity, wind speed, building deterioration(aging), floor space ratio, building coverage ratio and building use. The accuracy of prediction model using random forest algorithm was 74.7%.

**Key Words** : Data Mining, Electrical Fire, Machine Learning, Prediction Model, Spatial Data

### 1. 서론

재난안전에 대한 정부의 정책적 관심이 증대함으로 인하여 정부 주도 하에 재난안전 관련 다양한 사업들이 추진되기 시작했다. 국립안전재난연구원은 국민안전과 예방을 위해 국가가 보유한 필수적인 안전정보를 활용하여 생활안전지도를 구축하고, 이에 대한 대국민 서비스를 지속적으로 추진 중에 있으며, 행정안전부는 안전지수를 산출·공개함으로써 지자체별 안전수준을 확인하고 다양한 사업과 투자를 확대해 나갈 수 있는 정책을 지원하고 있다.

또한, 행정안전부는 전자정부지원사업을 통해 '17

년부터 이미 시범 서비스 중인 서민 생활 안전 강화 사업에 대해서 전국적인 서비스로 실시할 계획을 추진하고 있다. 이와 같이 다양한 재난안전의 정부 주도적 사업 추진 정책이 활발하게 이루어지면서 자연스럽게 재난안전에 대한 대국민적 관심과 새로운 기술의 니즈 역시 계속 증가했다. 재해·재난 분야의 경우, 시간이 지날수록 종류가 다양화되고, 규모가 커지면서 이에 대한 사전예측과 분석을 통해 안전을 담보 할 수 있는 데이터 기반 전략기획이 필요하게 되었다. 이는 곧 기존의 방어적 대응에서 벗어나 데이터마이닝 등 데이터 기반의 기획·검증 가능한 기술을 도입한 선제적 대응

\*Korea land and geospatial informatix corp., Chonbuk National University

\*\*Corresponding Author : Korea land and geospatial informatix corp. (tonyk7845@gamil.com)

Received December 10, 2018

Revised December 14, 2018

Accepted December 15, 2018

을 할 수 있는 기반의 필요를 의미한다.

그동안 정부는 매년 전기재해에 대한 사고유형 및 분석, 점검 등 전기재난 안전사고를 낮추기 위한 정책 등 급증하는 전기재해를 줄이기 위해 다양한 노력을 하였으나, 효율적인 의사결정지원 체계 및 기존 누적 데이터 활용방안의 미비로 효과적인 대처방안이 부재한 현황이다. 전기재해와 같이 특수성을 가진 재난 사고를 예방하기 위해서는 기존의 방대한 데이터 분석을 통한 데이터 기반 맞춤형 지원체계 및 대응체계가 필요하다.

한국전기안전공사는 매년 전기안전사고(화재·감전 등)에 대한 정량화된 통계 기반의 데이터를 수집하고 관리하면서, 데이터 기반의 전기재해를 예방하기 위한 전기안전점검관련 1억 2천만 건의 누적 데이터 활용 필요성을 검토하였다.

한편, 해마다 늘어나는 전력 수요에 따라 대형 전기 화재 피해액이 2배로 증가하였다. 최근 4개월 사이 대구 서문시장(‘16.11)과 인천 소래포구(‘17.03)등 대형 전기 화재 발생으로 인명·재산 피해가 대구 서문시장 전기화재 피해액은 약 469억 원, 인천 소래포구 전기화재 피해액은 6억 5,000만 원으로 추정된다. ‘09년부터 ‘15년까지 발생한 전기화재는 303,806 건으로 피해액 누적 4,602억 6,600만 원에 달하고 있으며, 특히 전기화재 피해액은 ‘06년 399억 원에서 ‘15년 723억 원으로 약 2배 가까이 급증하는 현황이다[1].

재래시장 전기화재의 경우, 전기시설 노후화와 전기 시설을 무분별하게 설치·사용하는 전기소비자의 안전불감증이 더해져 심각한 화재 원인으로 분석되고 있다. 이에 따라 인명 및 재산 피해가 급증하는 전기재해에 대응해 대국민 안전강화 및 재해 예방 사업추진 효율성을 제고할 수 있는 빅데이터 기반의 의사결정지원 체계가 촉구되었다.

이러한 전기적 화재사고에 의한 문제를 인지하고, 데이터 기반 데이터마ining 기술을 활용하여 전기안전 사고 예측을 통해 선제적 대응체계가 필요하다. 본 연구는 전기화재와 관련된 데이터 기반으로 화재사고 패턴을 파악하고, 반복적으로 발생하는 사고 유형과 사고 위험도를 기계학습을 통해 예측 및 검증함으로써 전기적 화재사고를 줄이는데 목적이 있다.

## 2. 연구방법

본 연구의 시간적 범위는 2012년 ~ 2016년 이고, 공간적범위는 대구광역시를 대상으로 진행하였다. 본 연구의 과정은 그림 1에 도식화 한 바와 같이 크게 데이터수집, 융합DB 구축, 알고리즘 개발, 검증 순으로 진행하였다. 데이터 수집단계에서는 기관별로 관리되고 있는 데이터를 수집하였다. 분석에서 핵심이 되는 화재사고데이터(119출동데이터), 한국전기안전공사에서 수집·관리되고 있는 전기안전점검데이터, 건축물속성데이터, 기상자료를 수집하였다. 다음으로 건물 중심으로 데이터를 융합(Join)하면서 융합 데이터셋을 구축하였다. 융합 데이터 구축과정에서 주소데이터, 좌표데이터를 공간화(geocoding)하고, 데이터 전처리를 수행 후 건축물정보(공간정보) 중심으로 융합하였다. 알고리즘 개발 단계는 전기화재를 예측모델로 만드는 과정이다. 전기화재 사고 유무데이터를 활용하여 화재사고가 발생하는 패턴을 도출하고 모델링을 수행하였다. 마지막으로 검증단계에서는 개발된 예측모델을 평가하는 단계로 학습에 사용하지 않은 테스트데이터를 통해서 정확도를 도출하고 실제사고와 예측된 사고의 유무와 비교 검증하였다.

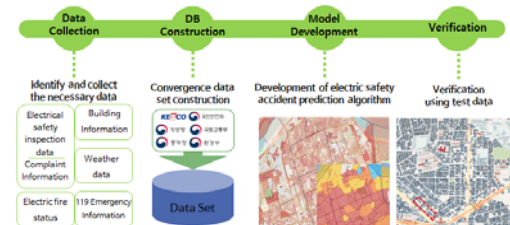


그림 1. 연구과정  
Fig. 1. Research Process

알고리즘 개발에서는 수많은 변수 중에서 종속변수에 영향을 끼치는 주요요인을 도출하였고, Random Forest 알고리즘 등 다양한 예측모델을 만들고 검증과정을 거쳐 정확도가 가장 높은 모델을 선정하였다. 분석결과를 통해 전기화재위험도를 점수화하고 등급화하였다. 등급화한 결과를 공간적으로 시각화하였고, 분석결과를 통해 전기안전 우선 점검지역 도출 등에 활용할 수 있도록 했다[2].

### 3. 데이터분석

#### 3.1. 데이터수집

원시데이터는 아래 <Table 1>과 같이 한국전기안전공사의 고객마스터 테이블, 점검결과 일 마감 테이블을 수집하였다. 또한, 국토교통부에서는 건축허가대장 테이블, 건물 등 별 개요 테이블, 건축허가대장 진행상황 테이블을 수집하였고, 행정안전부에서는 전기화재 현황 테이블을, 통계청은 행정동 경계 테이블, 인구통계 테이블, 기상청은 날씨정보 등 각각의 기관에서 요구 데이터를 수집하였다.

한국전기안전공사를 통해 확보한 총 1억 2천만 건 이상의 전기안전 점검결과 일 마감·고객마스터 데이터와 행정안전부에서 매년 파악한 3만 건 이상의 전기화재 현황 데이터를 기본으로 한다.

표 1. 데이터 목록

Table 1. Data List

| Data                    | Data field                                                              | quantity (records, volume) |
|-------------------------|-------------------------------------------------------------------------|----------------------------|
| Customer information    | date, first supply date, use_YN ...                                     | 2,521,782                  |
| Electrical safety check | customer no., check result, grade ...                                   | 120,212,444                |
| Building information    | usage, structure, completion date, coverage ratio, floor area ratio ... | 1,048,575                  |
| Total building GIS      | position, code, area ...                                                | 1,048,757                  |
| 119 fire accident data  | kind, causes, location, area ...                                        | 30,215                     |
| Weather data            | precipitation, humidity, temperatures, wind speed ...                   | 1.2G                       |
| Administrative district | administrative borders ...                                              | 70M                        |
| Emographics             | age, region ...                                                         | 30M                        |

#### 3.2. 데이터 전처리 및 융합

수집된 일반 속성정보 7종, 공간정보 2종의 데이터를

융합하기 위해 일반 속성정보를 공간화, 즉 지오코딩 과정을 거쳐서 모두 공간정보로 변환하고, 공간조인과정을 거쳐서 모든 수집된 데이터를 건물 단위로 융합하였다.

공간정보는 "지리적인 위치 및 위상 관계를 나타내는 도형정보와 공간상의 사물에 대한 특징을 나타내는 속성정보의 집합체"로 정의할 수 있다[3]. 도형정보는 지도학상의 좌표체계에 따라 좌표값을 기준으로 점, 선, 면 등의 형태로 표현되고, 속성정보는 아라비아 숫자나 텍스트 기호 등으로 표현된다. 특히, 공간정보를 입력할 때는 정보의 형상을 입력해야 하며 그 표현방법으로 벡터 데이터와 래스터 데이터의 형태가 있다. 공간정보의 가장 큰 특징은 도형정보와 속성정보가 1:1로 연결되어 관리된다는 것이다.

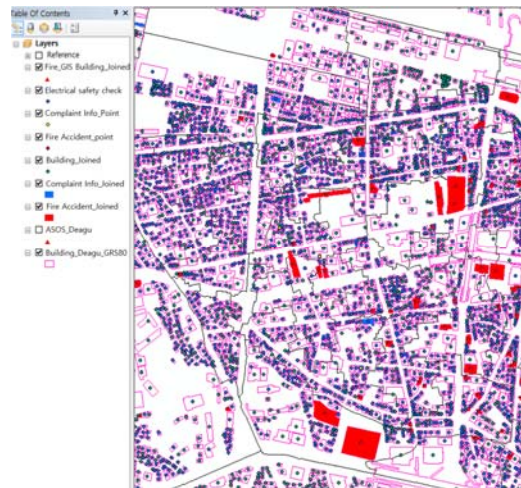


그림 2. 지오코딩(공간화)

Fig. 2. Geocoding

대구광역시에서 발생한 3천 여건의 전기화재사고를 중심으로 건물속성, 전기안전점검데이터, 기상청데이터를 위치와 사고일자 기준으로 융합하였다. 전기화재가 발생했던 건물의 속성과 점검결과데이터, 기온, 습도, 강수량 등 기상정보를 알아보려고 한다. 전기화재가 발생하지 않은 경우의 데이터셋을 만들기 위해 화재가 발생하지 않았던 건물공간정보 데이터에 전기안전점검데이터, 기상정보를 융합하였다.

최종 분석 데이터셋은 건물 중심으로 전기화재사고가 발생한 건물, 사고가 발생하지 않은 건물로 구분된

6,251건의(records) 데이터셋으로 만들었다. 속성정보는 건물별 전기화재사고유무, 건물 노후년수, 연면적, 건폐율, 용도, 구조 등 건축물 속성, 절연저항, 접지저항, 누설전류 등 전기안전점검정보, 기온, 습도, 강수량 등 기상정보건물 등 70개의 속성으로 구성했다.

### 3.3. 탐색적 분석

그림 3의 화재사고 데이터의 화재사고 장소 유형 분포를 살펴보면, 생활서비스 30.16%, 주거 27.58%, 산업시설 12.91%, 판매·업무시설 8.54%, 기타 4.57% 순으로 생활 서비스가 이루어지는 곳과 주거 지역에서 가장 많은 화재사고가 발생함을 알 수 있다.

또한, 화재사고 데이터의 주요 발화기기로는 배선/배선기구가 21.52%로 가장 많았으며, 전기설비 12.44%, 조명·간판 12.13% 등의 순으로 분포되어 주요 화재의 원인은 배선·배선기구 임을 알 수 있었다.

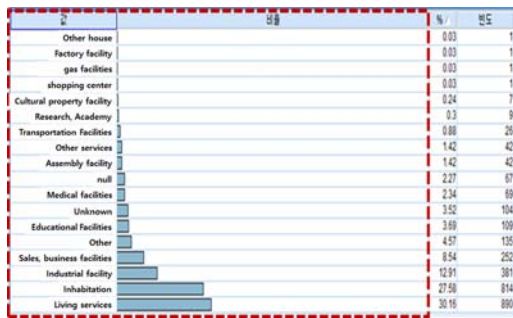


그림 3. 화재사고-발화기기 분포  
Fig. 3. Fire accident - Distribution of ignition equipment

화재사고의 발화요인 분포로는 압착 손상에 의한 단락이 59.84%로 대부분을 차지하고 있었다. 그 다음으로 전기적 요인 10.13%, 기타요인은 0.1% 순으로 나타났다.

또한, 주요 화재 유형으로는 건축, 구조물의 화재가 94.68%로 대부분을 차지하고 있으며, 쓰레기 화재 등 기타 화재 유형이 5.22%, 위험물 등에 의한 화재사고는 0.07%로 비율이 굉장히 낮은 것을 확인할 수 있다. 대부분의 전기적 화재는 건물에서 발생한다는 것을 알 수 있었다.

안전점검데이터의 업종분포로 분류하여 값을 살펴 보았을 때, 그림 4과 같이 주택이 54.53%로 가장 많았으며, 종교·학원기타 14.28%, 가로신호조명 10.48%로 그 뒤를 이었다.

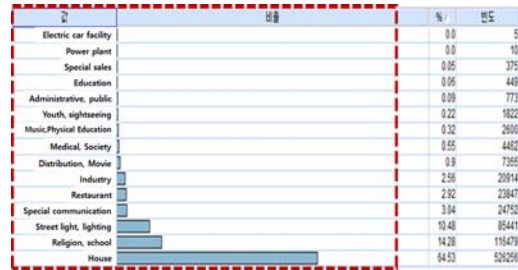


그림 4. 안전점검-업종분포  
Fig. 4. Safety inspection - category of business

RAW DATA 정제 결과에 따라 화재사고와 연관 인자별로 분석하였다. 먼저 화재사고와 전기적 요인(절연저항, IR)의 관계를 파악하기 위해 전기적 화재사고 시 절연저항 값과 무사고시 절연저항을 비교하였다.

그림 5에서 사고가 발생한 건물의 절연저항을 살펴 보면, 평균값이 84.5MΩ로 중앙값은 21.8MΩ로 나타났고, 사고가 발생하지 않은 건물의 절연저항의 평균값은 169.5MΩ, 중앙값은 65.3MΩ으로 나타났다. 이는 전기적 화재사고는 절연저항의 평균값이 낮을수록 사고 비율이 증가함을 알 수 있다.

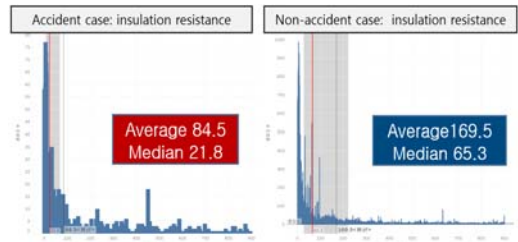


그림 5. 화재사고-전기적 요인(절연저항, IR)  
Fig. 5. Fire accident - Electrical factor (insulation resistance, IR)

또한, 화재사고 환경적 요인인 습도도 고려하여 데이터의 연관성 파악을 진행하였다. 그림 6에서 보는바와 같이 사고 습도의 경우 습도가 35~40% 일 때 20.79%의 사고율을 보였으며, 85~90%일 때 15.05%의 사고율이 도출되었다. 무사고 습도의 경우 고르게 분포하는 편이었으나 45~50%일 경우 9.017%의 무사고율이 도출되는 결과를 확인하였다. 이는 습도가 아예 높거나 낮을 때 사고가 높음을 확인할 수 있었다.

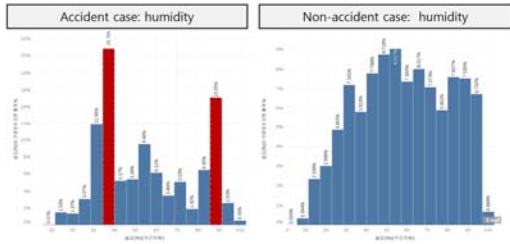


그림 6. 화재사고-환경적 요인(습도)  
Fig. 6. Fire accidents - Environmental factors (humidity)

### 3.4. 예측 모델링

전기화재사고 유무에 대해서 전기적, 환경적 요인과 관련된 주요요인을 찾고, 예측모델을 통해서 건물단위 화재사고를 예측하고 위험도를 계량화하는 단계이다. 이를 위해서는 기존 관리되어왔던 전기안전점검 데이터, 건축물 대장, 화재사고 데이터 등을 활용해 다양한 요인을 구분하고 사고가 많이 발생한 지역과 아닌 지역 간의 특성을 찾아냈다.

융합된 데이터셋 중 60%를 학습데이터로 활용하여 Random Forest 알고리즘을 통해 학습모델을 만들었다. L.Breiman에 의해 제안된 Random Forest는 다수의 결정 이진(binary) 트리를 앙상블 형태로 결합한 것으로, 각 이진트리에서는 랜덤한 방법으로 트리들을 성장 시킨다[4]. Random Forest는 결정 트리들을 기본으로 하고 있기 때문에, 빠른 학습속도와 많은 양의 데이터 처리 능력을 가지고 있다[5]. 즉, Random Forest는 여러 개의 Decision Tree를 만들고, voting을 통해서 최종 결과를 도출하는 방법이다. 데이터를 bootstrap aggregating(bagging)과정을 통해 N개의 샘플링 데이터셋(observations과 features들을 random하게 sampling)을 구성한 후 각각의 Decision Tree 모델을 구성하는 한 후 개별 예측모형이 voting 방식으로 예측결과를 결정하여 Low Bias는 유지하고 High Variance는 줄이는 방법이다[6]. Random Forest는 이런 배경 계열의 가장 대표적이고 예측력이 좋은 알고리즘이다. 예측결과와 정확성(Low Bias)은 개별 예측모형에 쓰이는 알고리즘(decision tree)의 평균값으로 유지되는 반면 낮은 안정성(High Variance)은 Central Limit Theorem에 의해 낮아진다[7]. 융합 데이터 셋을 Random forest 알고리즘

을 활용하여 예측모델 개발하였다. 그 결과, 평균절대 오차(Mean Absolute Error)는 0.261(건), 실제사고와 예측된 사고와 상관계수가 0.725로 나타났다. 평균절대오차는 오차의 절대치, 즉, 절대편차를 모두 더한 다음 이를 레코드수로 나눈 값을 의미한다[8]. 다음 식은 예측의 정확도를 평가하기 위한 MAE의 계산식이다.

$$MAE = \frac{1}{N} \sum_{j=1}^N |R_j - \widehat{R}_j| \dots\dots\dots (1)$$

여기서,  $R_j$  는 실제 사고유무 값이고,  $\widehat{R}_j$  은 예측된 사고유무 값이다. 상관분석은 두 변수 간에 얼마나 밀접한 선형관계를 가지고 있는 가를 분석하는 통계기법으로, 두 변수간의 관계의 강도를 상관관계(correlation analysis)라 한다. 상관관계는 분석방법에 따라 단순히 두 개의 변수가 어느 정도 유사한가를 측정하는 기법이다[9]. 상관관계가  $0 < r \leq +1$  이면 양의 상관,  $-1 \leq r < 0$  이면 음의 상관,  $r = 0$ 이면 무상관이라고 한다[10]. 그러나 0인 경우 상관이 없다는 것이 아니라 선형의 상관관계가 아니라는 것이다. 본 연구에서는 이와 같은 이론을 근거로 사용되고 있는 피어슨 상관계수(Pearson correlation coefficient)를 활용하여 실제 화재사고와 예측된 화재사고와 유사성을 살펴보기 위해 상관분석을 실시하였다. 피어슨 상관 계수는 다음과 같이 정의한다.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

여기서  $r$  은 실제사고와 예측사고와의 유사도 가중치이며,  $x_i$  는 실제 사고유무를 나타내고,  $y_i$  는 예측된 사고유무를 나타낸다.

그림 7에서 보는바와 같이 주요 예측변수로는 습도, 풍속, 건축물 노후년수, 용적율, 건폐율, 절연저항(IR), 건축물용도, 건축물 구조로 나타났다.

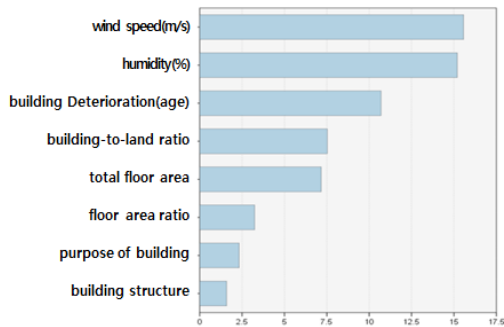


그림 7. 예측변수 중요도  
Fig. 7. Geocoding

### 3.5. 검증

예측모델을 검증하기 위해 모델링에 활용하지 않은 40%의 테스트데이터를 활용하여 그림 8과 같이 예측 모델을 검증하였다.

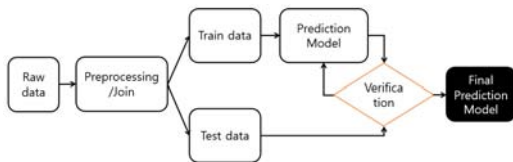


그림 8. 알고리즘 개발 흐름도  
Fig. 8. Algorithm Development Flow

최종적으로 개발된 전기적 요인 및 환경적 요인을 활용한 예측 모델을 검증하기 위해 검증데이터를 통해 예측된 결과값을 도출하였다. 그림 9의 예측값이 0.5 이상이면 화재 발생(1), 0.5 이하면 화재 미발생(0)으로 구분하여 실제 사고 유무와 교차표(Matrix chart)를 만들었다. 예측모델을 통하여 건물 단위로 사고를 사고로 예측하는 정확도가 74.7%로 확인되었다.

| Fire acc_YN |          | 0      | 1      |
|-------------|----------|--------|--------|
| 0           | count    | 12365  | 1642   |
|             | column % | 88.277 | 11.723 |
|             | total %  | 44.137 | 5.861  |
| 1           | count    | 3543   | 10465  |
|             | column % | 25.293 | 74.707 |
|             | total %  | 12.647 | 37.355 |

그림 9. 검증결과  
Fig. 9. Verification Result

예측모델에 적용하여 건물별 위험도를 도출하여 지도로 시각화해보면 그림 10과 같다. 사고 난 곳을 위험

하다고 나타내는 경우도 있지만, 사고는 나지 않았지만, 위험도가 높게 예측한 건물을 주의깊게 점검할 필요가 있다고 판단된다.



그림 10. 예측 결과와 실제 사고 비교  
Fig. 10. Comparing forecast results with actual incidents

최종 예측모델을 대구지역 22만개 건물데이터에 적용하여 분석결과를 도출하였다. 전기화재위험도 예측 모델의 핵심변수인 습도, 풍속, 건물 노후년수, 용적률, 건폐율 IR Value, 건축물용도, 건축물구조 등을 통해 대구지역 안전, 관심, 주의, 경계, 위험 등급을 그룹화하여 Range별로 분류하고, 분류된 Range별 건수를 확인하여 1등급에서 5등급까지 각 비율을 비교하였다.

예측결과 데이터를 Jenks Natural Breaks Classification을 활용하여 건물별 위험도를 5가지 등급(위험, 경계, 주의, 관심, 안전)으로 분류하고 공간조인을 통해 시각화하였다. Jenks Natural Breaks Classification은 같은 등급 내 전체 값들의 평균을 기준으로 평균편차는 최소화되고, 각 등급 간의 분산은 극대화하는 것이다. 즉, 등급 내의 분산은 줄이고 등급 간의 분산을 최대화하는 방법이다[11].

그림 11에서 0.015000 ~ 0.147000 값인 197,439건에 대해 1등급 71.7%로 안전한 것으로 도출되었으며, 0.147001 ~ 0.276000 값인 31,019건에 대해 2등급 11.3%로 관심단계, 0.276001 ~ 0.409000 값인 19,863건에 대해 3등급 7.2%로 주의 단계, 0.409001 ~ 0.546000 값인 20,008건에 대해 4등급 7.3%로 경계단계, 0.546001 ~ 0.857000 값인 6,955건에 대해 5등급 2.5%로 위험단계로 나타났다.

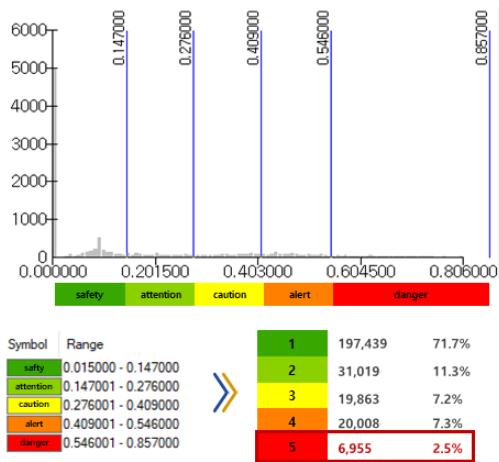


그림 11. 등급별 범위 및 건수  
 Fig. 11. Range by grade and count

#### 4. 결 론

본 연구는 환경적·전기적 요인에 따른 복합적인 이유에서 발생하는 전기화재사고를 데이터기반의 머신러닝 기술을 적용하여 사고를 예측하고 이를 활용하여 사전에 예방점검을 통해 전기화재발생률을 줄이기 위함이다.

본 연구를 통해 도출된 결과를 살펴보면, 첫째, 내부 데이터인 한국전기안전공사의 점검결과 일마감 데이터, 고객마스터 데이터를 기반으로 국토교통부, 기상청, 통계청 등 외부 필요 데이터 수집 및 융합 관리할 수 있는 데이터베이스를 구축하였다. 둘째, 건축단위 화재사고 위험도를 예측하고, 건축물, 행정구역 등 공간데이터셋을 융합하여 GIS기반의 전기안전지도 서비스를 구현함으로써 실시간 전기화재 예측에 기여하였다. 마지막으로 사고유무 속성을 종속변수로 하고 습도, 풍속, 건축물 노후년수, 용적율, 건폐율, 절연저항(IR), 건축물용도, 건축물 구조 등을 설명변수로 설정하여 예측모델을 개발하였다. Random Forest 알고리즘을 활용하여 개발된 예측모델을 검증한 결과 74.7%의 높은 예측정확도를 확보하였다. 이 값은 2017년 행정안전부의 빅데이터 활용 화재위험도 예측 시스템 구축 보고서의 예측정확도 64.02%와 비교해볼 때 알고리즘의 예측정확도가 10.68% 높았다[12].

향후 전국 건축물을 대상으로 딥러닝(deep learning) 기반 예측모델로 고도화하여 정확도를 개선하는 연구가

필요하다. 이어서 본 연구결과를 토대로 IoT 실시간 데이터 연계, 분석을 통해 전기화재 예방 및 감소를 위한 연구를 진행할 예정이다.

#### REFERENCES

- [1] S. K. Lee, "A statistical Analysis on the electrical Accident", KESCO, Aug. 2016.
- [2] Breiman, L. Bagging predictors, Machine Learning, 24, 123-140. 1996
- [3] J. S. Ryu, S. T. Kim, "A conceptual Framework for spatial database structure", Korea Research Institute for Human Settlements., May 1996.
- [4] L. Breiman. "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, Oct. 2001
- [5] B. C. Ko, S. H. Kim, and J. Y. Nam, "X-ray image classification using random forests with local wavelet-based CS-local binary patterns," J. Digit. Imaging, vol. 24, no. 16, pp. 1141-1151, Oct. 2011
- [6] J. H. Hong, B. C. Ko, J. Y. Nam, "Human Action Recognition in Still Image Using Weighted Bag-of-Features and Ensemble Decision Trees", The Journal of Korea Information and Communications Society, Vol.38A(1), pp.1-9, Jan 2013.
- [7] H. Jung, and J. Kim, "A Machine Learning Approach for Mechanical Motor Fault Diagnosis," Society of Korea Indus. and Sys. Eng., vol. 40, no. 1, pp. 57-64, 2017.
- [8] F. Robert Jacobs, "Operations and Supply Management", McGrawHill, 2010
- [9] Y. T. Hur, D. I. Shin, S. K. Lee, "Data Mining of Gas Accident and Meteorological Data in Korea for a Prediction Model of Gas Accidents", Journal of the Korean Institute of Gas, Vol. 16, No. 1, Feb 2012.
- [10] D. S. Bea, "Statistical theory and application", chungmungak, 2003.
- [11] Jenks, George F., "The Data Model Concept in Statistical Mapping", International Yearbook of Cartography 7, pp. 186-190,

1967.

[12] Ministry of Public Administration and Security, "Development of a Fire Risk Prediction System using Big Data", 2017.

[13] Jae-Young Chang, 'An Experimental Evaluation of Box office Revenue Prediction through Social Bigdata Analysis and Machine Learning', The Journal of The Institute of Internet, Broadcasting and Communication VOL. 17 No. 3, 2017

---

저자약력

---

**고 경 석(Kyeong-Seok Ko)** [정회원]



- 서울도시철도공사  
System Engineer
- 한국국토정보공사  
System Manager
- 전북대학교 박사과정

<관심분야> 머신러닝, 인공지능, 패턴인식, 데이터분석

**황 동 현(Dong-Hyun Hwang)** [정회원]



- 모바일웬컴 대표이사
- 한국국토정보공사  
융복합사업부 팀장
- 주식회사 업데이트 대표

<관심분야> 빅데이터, 인공지능, 예측 서비스 등

**박 상 준(Sang-June Park)** [정회원]



- 전북대학교 경영학과 교수
- 전북대학교 상과대학 학장
- KAIST 경영과학 박사

<관심분야> 마케팅모델링, 마케팅조사, 수요예측 등

**문가경(Ga-Gyeong Moon)** [정회원]



- 연유아카데미 대표
- 한국전기안전공사  
콜센터 센터장
- 전북대학교 박사 수료

<관심분야> 빅데이터, 인공지능, 예측 서비스, 마케팅조사, 고객만족분석 등