



차세대융합기술학회논문지 제6권 8호

ISSN : 2508-8270(Print)

기상 정보와 소방 데이터를 연계한 강원도 지역의 화재 발생 가능성 예측 연구

김봉현

To cite this article : 김봉현 (2022) 기상 정보와 소방 데이터를 연계한 강원도 지역의 화재 발생 가능성 예측 연구 , 차세대융합기술학회논문지, 6:8, 1337-1346

① earticle에서 제공하는 모든 저작물의 저작권은 원저작자에게 있으며, 학술교육원은 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

② earticle에서 제공하는 콘텐츠를 무단 복제, 전송, 배포, 기타 저작권법에 위반되는 방법으로 이용할 경우, 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

www.earticle.net

기상 정보와 소방 데이터를 연계한 강원도 지역의 화재 발생 가능성 예측 연구

김봉현*

*서원대학교 소프트웨어학부 컴퓨터공학과 교수

A Study on Prediction the Possibility of Fire in Gangwon-do by Linking Weather Information and Firefighting Data

Bong-Hyun Kim*

*Professor, School of Software, Department of Computer Engineering, Seowon University

요약 최근 10년간 국내 화재발생 건수 및 인명 피해는 일정한 수준을 유지하고 있다. 화재는 다양한 원인에 의해 발생하지만, 기상 정보의 영향을 많이 받는다. 특히, 우리나라의 경우는 고온다습한 여름에 화재 발생 건수가 감소하고, 저온저습한 겨울철에는 화재 발생 건수가 증가하는 추세이다. 화재는 언제 어디서든지 발생할 수 있는 것으로 막대한 피해를 초래할 수 있기 때문에, 이에 대한 예측을 통해 피해를 최소화시킬 필요가 있다. 따라서, 본 논문에서는 화재 사고 원인 분석, 계절별 화재 발생률 등을 분석하고 기온, 풍향, 풍속, 습도와 같은 기상 정보와 화재 발생 건수와의 연관성을 분석하였다. 이를 통해, 기상 정보 입력으로 화재 발생 가능성을 예측하는 모델을 설계하였다. 최종적으로, 4개의 독립 변수(기온, 풍향, 풍속, 습도) 값을 입력하면 변수에 저장한 뒤 회귀 모델에 적용하여 최종 예측 모델을 설계하였다.

주제어 : 화재예측, 기상정보, 소방데이터, 사이킷런, 지오코더

Abstract Over the past 10 years, the number of domestic fires and casualties has been maintained at a constant level. Although fires are caused by various causes, they are greatly influenced by weather information. In particular, in Korea, the number of fires decreases in hot and humid summer, and the number of fires increases in winter when low and low humidity. Since a fire can occur anywhere at any time and can cause enormous damage, it is necessary to minimize the damage through prediction. Therefore, in this paper, the cause of fire accidents and the rate of fire occurrence by season were analyzed. In addition, the correlation between weather information such as temperature, wind speed, wind direction, and humidity and the number of fires was analyzed. Through this, a model was designed to predict the possibility of fire by inputting meteorological information. Finally, when the values of four independent variables (temperature, wind direction, wind speed, humidity) are input, they are stored in the variables and then applied to the regression model to design the final predictive model.

Key Words : Fire prediction, Weather information, Firefighting data, Sklearn, Geocoder

* Corresponding Author : Bong-Hyun Kim (bhkim@seowon.ac.kr)

Received August 12, 2022

Revised August 17, 2022

Accepted August 22, 2022

Published August 28, 2022

1. 서론

화재는 불로 인한 연소 현상으로 다양한 원인으로 인해 발생한다. 화재는 발생 대상에 따라서 건축물에 발생하는 건물화재, 산림 또는 논, 밭에 발생하는 임야 화재, 자동차 또는 이륜차 등에 발생하는 차량화재, 선박에 발생하는 선박화재, 비행기 등에 발생하는 항공기화재, 기타 화재의 여섯 종류로 구분된다. 또한, 원인에 따라 분류하면 방화, 실화, 자연발화, 낙뢰, 기타 원인 등 다섯 종류로 구분되고, 소실 정도에 따르면 전소, 반소, 부분 연소 등 세 종류로 구분할 수 있다. 이러한 화재로 인해 막대한 인적, 물적 피해를 발생하고 있는 실정이다[1-3].

우리나라의 경우, 최근 10년간 평균 화재 건수는 42,332건, 인명피해 2,215명, 재산 피해 476,529백만 원으로 나타났으며, 도시화로 인한 인구 및 산업 시설이 증가함에 따라 화재로 인한 인명 및 재산 피해는 증가하고 있는 추세이다[1][2]. Fig. 1은 2011년부터 2020년까지의 국내 화재 건수 및 인명 피해 현황을 나타낸 것이다.



Fig. 1. Current status of domestic fires and casualties

화재는 다양한 원인이 있지만, 기상 상태에 따라 화재 발생 및 피해 규모의 차이가 발생한다. 특히, 우리나라는 연중 기상 상태가 달라지기 때문에 1년 중 화재가 많이 발생하는 것은 겨울철 12, 1, 2월경이다. 이는 그 계절에 불의 사용도가 높고 취급의 부주의로 인한 것으로 분석되고 있다. 화재가 다발하는 계절은 지역에 따라서도 다르나, 대개 기온이 낮은 겨울 또는 봄에 많으며, 이것은 습도, 계절풍에도 다소 관계가 있다. 또한, 화재와 상대습도는 매우 밀접한 관계가 있다. 건조한 날에는 화재 발생수가 많고, 습한 날에는 화재 발

생수가 적다[4]. 이는 습도가 낮으면 발화시 연소 속도가 커지므로 발화되기 쉽기 때문이다. 계절별 화재 발생건수를 보면 겨울과 봄이 비슷한 경향을 보이고 여름과 가을이 거의 같은 경향을 보인다. 여름과 가을은 같은 상대습도 값에 대하여 겨울이나 봄에 비해 화재 건수도 2/3정도 적는데 이러한 차이는 추운 겨울과 봄철에 불의 사용빈도가 많기 때문이다. 특히 겨울과 봄철에는 대기 습도가 30~40%밖에 안 되고, 또한 이때는 공기가 압축되어 공기 속에 산소 함유량이 많아짐으로서 연소 속도도 증가하기 때문이다. 보통 연소는 기온이 높을 때 활발하고, 기온이 낮을 때에는 활발하지 않다. 겨울과 봄철에는 다 같이 습도가 낮아 화재발생이 염려가 크지만, 겨울에는 기온이 낮아 봄에 비해 선 발화가 어려운 편이다. 큰 화재의 경우를 보면 단순히 습도가 높고 낮음보다 오히려 얼마나 오랫동안 건조한 날씨가 계속되었는가에 의해 더 큰 영향을 받는다는 것을 알 수 있다[5].

또한, 화재 발생과 풍향과는 그리 큰 관계가 없는 것으로 알려져 있다. 하지만 규모가 큰 화재의 경우는 바람이 강한 날에 일어나는 비율이 높는데, 이는 일단 불이 나면 불티가 바람을 타고 광범위하게 급속도로 퍼질 뿐만 아니라 연소도 극히 빨라지기 때문이다. 이와 같이, 화재가 발생하기 쉬운 기상조건은 오랫동안 대기가 건조한 상태에서 바람이 강하며 습도가 낮을 때인 것으로 파악할 수 있다.

이러한 화재 피해를 줄이기 위해 예방, 대비, 대응, 복구의 단계로 재난 관리를 하고 있다. 이 중 화재 예방은 화재 발생 가능성을 감소시키거나, 화재가 발생한 경우 그 영향을 제한하기 위한 활동을 의미한다. 화재 예방 단계는 화재의 발생을 근본적으로 제어한다는 점에서 화재 피해를 최소화하기 위한 가장 중요한 단계이다[6]. 이러한 중요성으로 인해 화재 재난 관리의 패러다임은 화재 대응에서 화재 예방으로 변화하고 있으며, 전 세계적으로 다양한 화재 예방 계획이 도입되고 있다.

이러한 추세를 반영하여, 화재 발생 이전에 이를 예측하고 사전에 피해를 최소화시키려는 노력이 반드시 필요하다. 이를 위해, 최근에는 미래 화재 안전 기술의 개발 및 활용 분야에서 정보통신기술(ICT)과 빅데이터 분석 기술이 빠르게 발전함에 따라 기존의 과거 통계

중심의 데이터 분석이 아닌 빅데이터 분석 방법으로 과학적 화재 사고 예방 관리 시스템 구축과 이를 위한 화재 피해 예측 모델 개발의 필요성을 강조하고 있다.

따라서, 본 논문에서는 화재 사고 원인 분석, 계절별 화재 발생률 등을 분석하고 기온, 풍향, 풍속, 습도와 같은 기상 정보와 화재 발생 건수와의 연관성을 분석하였다. 이를 통해, 기상 정보 입력으로 화재 발생 가능성을 예측하는 모델을 설계하였다.

II. 기존 연구 분석

화재 피해 및 사고 예측 방법으로 일반적으로 많이 사용되는 회귀분석 방법이 있다. 빅데이터 분석 방법 중 하나인 회귀분석은 독립변수가 종속변수에 영향을 미치는지, 변수 사이의 관계를 알아보고자 할 때 실시하는 분석 방법이다. 회귀분석은 독립변수와 종속변수가 서로 상관관계를 가질 때 유의미한 결과 혹은 예측에 높은 설명력을 가질 수 있으며 독립변수가 변화함에 따라서 종속변수가 어떻게 변화하는지에 대해 규명하는 통계 분석 방법이다[7-9].

또한, 회귀분석으로 독립변수에 따른 종속변수의 관계를 선형 방정식 도출을 통하여 종속변수의 추정 및 예측이 가능하다. 회귀분석은 크게 단순회귀분석과 다중회귀분석 두 가지로 나누어지는데, 하나의 독립변수로 종속변수와의 관계를 분석하는 것이 단순회귀분석이고, 두 개 이상의 독립변수들과 하나의 종속변수와의 관계를 분석하는 것이 다중회귀분석이다. 이러한 회귀분석으로 독립변수에 따른 종속변수의 관계를 선형 방정식 도출을 통하여 종속변수의 추정 및 예측이 가능하다. 회귀분석을 통하여 금융 관련 예측, 판매 예측, 시계열 예측 등에 응용하여 넓게 활용되고 있으나 재난 및 화재 분야에 활용한 사례는 부족한 실정이다[10][11].

따라서 회귀분석 방법으로 화재 피해에 영향을 미치는 요인에 따른 화재 피해 및 발생 예측에 관한 필요성이 부각되고 있다.

대표적인 기존 연구 사례를 살펴보면, Shai는 다중 회귀분석을 통해 미국 필라델피아의 인구학적 요인(저소득, 고연령, 빈 주택 등)과 화재 부상률과의 관계를

조사하였다. 그 결과 빈 주택 비율, 저소득 및 비영리권 인구요인이 화재 부상률에 영향을 미치는 것으로 나타났다. Gilbert는 화재 사망 및 부상의 가능성에 영향을 미치는 사회 인구학적 요인(인구수, 성별, 인종, 연령, 빈 주택, 실업률 등)을 조사하였으며, 인구통계학적 정보를 사용하여 화재 건수를 예측하는 다양한 회귀 모델을 제시하였다. M.Taylor는 인구 특성 변화와 화재 위험(화재 건수, 인명피해, 재산피해)의 통계적 분석을 통해 영국 Merseyside 지역의 인구 변화 추세에 따른 화재 위험 결과를 분석하였다. 그 결과 노인 및 장기적인 건강 문제, 장애가 있는 사람들과 관련된 화재 위험이 증가할 가능성이 있고, 흡연 및 알코올 소비와 관련된 화재 위험이 감소할 가능성이 있다는 사실을 밝혔다[12-14].

또한, Volkan Sevinc는 베이지안 네트워크 모델을 사용하여 산불에 영향을 미치는 요인(상대 습도, 온도, 풍향, 화재 발생 지점과 정착지까지의 거리, 화재 면적, 화재 발생 지점과 농경지까지의 거리, 화재 발생 지점과 도로까지의 거리)에 따른 산불 발생 예측 모델을 개발하였으며, 온도가 산불 발생에 가장 큰 영향을 미치는 요인이라는 결론을 도출하였다. Kajol R. Singh는 기계학습 분석 방법 중 하나인 support vector machine 모델을 사용하여 인도 산림 지역의 기상 데이터 속성 요인에 따른 산불 화재 발생 예측 모델을 개발하였다[15][16].

국내에서는 공공데이터를 활용하여 건물 데이터와 화재 피해 데이터를 기반으로 인공 신경망 모델을 통해 건축물 특성(대지면적, 건축물 면적, 건폐율, 용적률, 층수, 건물 연령)에 따른 건물 단위 화재 발생 위험도를 예측할 수 있는 모델을 개발하였다. 또한, 전기안전 점검 데이터와, 전기 화재 사고 정보, 건축물 정보, 기상청 데이터를 기반으로 랜덤 포레스트 알고리즘을 활용하여 절연저항 값, 습도, 풍향, 건축물 노후도, 용적률, 건폐율에 따른 전기 화재 발생 예측 모델을 개발하였다[17-19].

III. 데이터 수집 및 관리

3.1 데이터 수집

본 논문에서는 강원도 지역의 화재 사고 원인 분석

및 계절별 화재 발생률 등을 분석하기 위해 소방안전 빅데이터 플랫폼과 기상청 빅데이터 정보를 활용하여 데이터를 수집하였다. 이를 위해, 2020년 1월 1일부터 2021년 6월 30일까지 강원도 지역에서 발생한 화재에 대한 데이터 2160개(결측치 제거)를 대상으로 데이터 수집, 정제, 분석, 예측을 진행하였다. 수집한 데이터는 CSV 파일로 불러와 fire_df 객체에 저장하였으며, fire_df 객체에 저장된 내용을 위에서부터 5개 행만 출력해 오류 여부를 확인하였다.

또한, 데이터 정제는 필요한 데이터를 원인/계절별 화재 발생률 분석을 위한 데이터 컬럼, 화재 발생 장소 표시를 위한 위치 데이터 컬럼, 기상정보와 화재 연관성 분석을 위한 데이터 컬럼으로 정리하였다. 아래 Table 1은 각 데이터 컬럼별 정보를 나타낸 것이다.

fire_df 객체에서 원하는 데이터를 정제하여 fdf 객체에 저장하였으며, 분석 시 오류를 방지하기 위해 결측치(NaN 값)를 제거하였다. 결측치 제거를 위해, dropna() 함수를 활용해 결측값이 있는 행을 제거하였으며, 제거 확인을 위해 fdf 객체에 저장된 내용을 출력하여 확인하였다. 이 과정에서, 일부 데이터가 제거되었으며 이러한 결과는 위치 컬럼(LOC_INFO_X와 LOC_INFO_Y)에 결측치가 존재했기 때문에 발생한 현상으로 LOC_INFO_X에 결측치가 존재하는지 isnull() 함수를 활용해 확인하였다.

Table 1. Data column by information

Data column for fire rate analysis by cause and season			
MSFRTN_RESC_R EPRT_NO	ACDNT_CAUSE	SEASON_SE_NM	
Disaster relief report number	Cause of accident	Season Classification	
Location data column to indicate where a fire occurred			
LOC_INFO_X		LOC_INFO_Y	
Location information X		Location information Y	
Data column for analyzing the correlation between weather information and fire			
TIME_UNIT_ TMPRT	TIME_UNIT_ _WS	TIME_UNI T_WD	TIME_UNIT_ HUMIDITY
Time unit temperature	Hourly wind speed	Hourly wind direction	Hourly unit humidity

3.2 데이터 관리

데이터 분석 및 예측을 위한 관리는 7개 영역으로 구분하여 진행하였다. 먼저, 화재 사고 원인 분석 영역에서는 작동기기, 불꽃·불티, 담뱃불·라이터불, 미상, 마찰·전도·복사, 기타, 화학적 발화열, 폭발물·폭죽, 자연적 발화열의 총 9가지 원인이 존재하는 것으로 파악하였다. 둘째, 계절별 화재 발생률 영역에서는 전체 화재 데이터 중 각 계절인 봄, 여름, 가을, 겨울에 몇 번의 화재가 발생했는지를 파악하였다. 셋째, 기온별 화재 발생 현황 영역에서는 영하 22°C부터 영상 36°C 사이의 기온 데이터를 적절한 범위로 나눠 화재 발생 건수를 측정하였다. 넷째, 풍향·풍속별 화재 발생 현황 영역에서는 풍향·풍속 데이터를 활용해 화재 발생 건수를 측정하였다. 다섯째, 습도별 화재 발생 현황 영역에서는 0%부터 100% 사이의 습도 데이터를 적절한 범위로 나눠 화재 발생 건수를 측정하였다. 여섯째, 기상정보의 영향을 받은 화재의 발생 위치 영역에서는 화재 발생 건수가 많았던 기온, 풍향, 풍속, 습도의 조건을 모두 만족하는 화재의 발생 위치를 발생 당시 계절과 함께 표시하고 계절별 발생 건수를 시각화하였다. 마지막으로, 산점도와 선형 회귀 그래프 영역에서는 화재 발생 건수와 기온, 풍향, 풍속, 습도의 상대적인 빈도를 측정해 회귀 분석 모델을 구축하였고, 구축한 모델을 바탕으로 기상정보를 입력해 화재 발생 가능성 예측을 수행하였다.

IV. 분석 및 예측

4.1 데이터 분석

본 논문에서는 화재 사고 원인 분석, 계절별 화재 발생률 등을 분석하고 기온, 풍향, 풍속, 습도와 같은 기상 정보와 화재 발생 건수와의 연관성을 분석하였다. 이를 위해, 화재 사고 원인을 분석하여 파이 차트로 시각화하였다. 사고 원인(ACDNT_CAUSE) 컬럼으로 그룹화한 후에 ac_group_sr을 바탕으로 파이 차트를 그려 출력하였다. Fig. 2는 화재 사고 원인을 분석한 출력 차트를 나타낸 것이다.

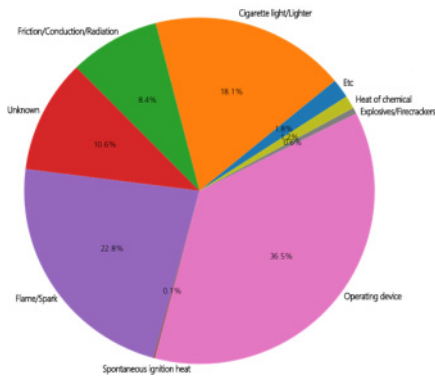


Fig. 2. Results of fire accident cause analysis

화재 사고의 9개 원인을 파이 차트로 시각화하였으며, 작동 기기로 인한 화재 발생이 가장 큰 비율을 차지하였고, 불꽃·불티로 인한 화재 발생이 두 번째로 큰 비율을 차지함을 알 수 있다.

계절별 화재 발생률 파이 차트 시각화에서는 계절 구분명(SEASON_SE_NM) 컬럼으로 그룹화한 후 파이 차트를 출력하였다. 각 계절에 해당하는 화재 발생 건수를 확인하고, ssn_group_sr을 바탕으로 파이 차트를 그려 출력하였다. Fig. 3은 계절별 화재 발생률을 분석한 출력 차트를 나타낸 것이다. 화재 사고가 발생한 계절의 비율을 파이 차트로 시각화하였으며, 화재 발생률은 봄, 겨울, 여름, 가을 순으로 높게 분석되었다.

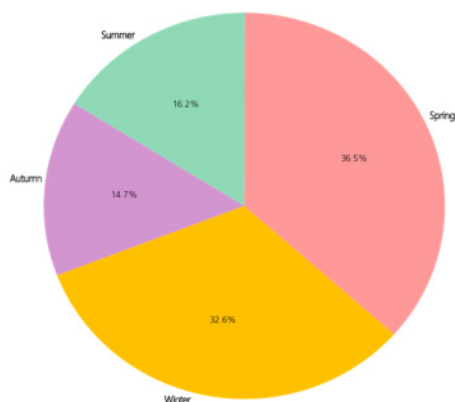


Fig. 3. Fires rate by season

기온별 화재 발생 현황 막대그래프 시각화에서는

단위 시간 기온(TIME_UNIT_TMPRT) 컬럼으로 그룹화하고, fdf 객체에 기온별 화재 발생 건수 구분을 위한 태그(TMP_TAG) 컬럼을 추가하였다. 막대그래프 X축에 기온의 범위를 활용하기 위해 TMP_TAG 컬럼을 추가하였으며, fdf 객체에 저장된 내용을 출력하여 추가된 TMP_TAG 컬럼을 확인하였다. 또한, 기온 범위를 표현하기 위한 리스트 생성과 함수(find_TMP)를 정의하였다. tmp_label 리스트는 막대그래프의 X축에 해당하는 레이블이며, tmp_count 리스트는 groups 리스트의 각 원소가 몇 개인지 세는 리스트이다. groups 리스트는 임의로 설정한 기온의 범위를 구분하는 태그이며, find_TMP 함수는 value 값으로 행의 TIME_UNIT_TMPRT 값을 받은 뒤, 임의로 설정한 기온 범위에 해당하면 tmp_count 원소에 1을 더하고 groups 리스트의 원소를 반환하였다. 만약 8°C라면 0°C 이상 10°C 이하의 범위에 속하기 때문에 TMP_TAG 값은 10이 되게 설정하였다. 또한, find_TMP 함수를 활용해 기온별 화재 발생 건수를 측정하였다. TMP_TAG 컬럼에 find_TMP 함수에서 반환된 값을 저장하고, tmp_count 리스트를 출력해 기온별 태그에 해당하는 화재 발생 건수를 출력하였다. 마지막으로, X축은 기온(°C), Y축은 화재 발생 건수인 막대그래프를 그려 출력하였다. Fig. 4는 기온별 화재 발생 현황을 나타낸 것이다. 분석 결과, 기온이 1°C 이상 20°C 이하인 상황에서 화재 발생 가능성이 높은 것으로 도출되었다.

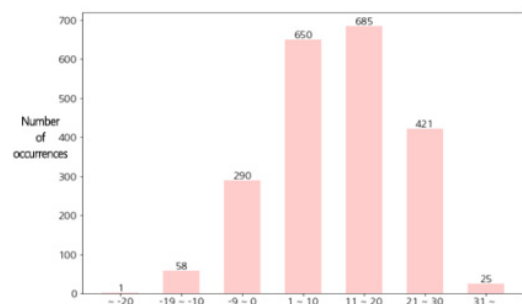


Fig. 4. Fires by temperature

풍향별 화재 발생 현황 막대그래프 시각화에서는 단위 시간 풍향(TIME_UNIT_WS) 컬럼으로 그룹화

하였다. fdf 객체에 풍향별 화재 발생 건수 구분을 위한 태그(WS_TAG) 컬럼 추가하고, 막대그래프 X축에 풍향을 활용하기 위해 WS_TAG 컬럼을 추가하였으며, fdf 객체에 저장된 내용을 출력해 추가된 WS_TAG 컬럼을 확인하였다. 또한, 풍향을 표현하기 위한 리스트 생성과 함수(find_WS)를 정의하였다. ws_label 리스트는 막대그래프의 X축에 해당하는 레이블이며, ws_count 리스트는 groups 리스트의 각 원소가 몇 개인지 세는 리스트이다. groups 리스트는 임의로 각각의 풍향을 구분하는 태그이며, find_WS 함수는 value 값으로 행의 TIME_UNIT_WS 값을 받은 뒤, 풍향을 구분하는 태그에 해당하면 ws_count 원소에 1을 더하고 groups 리스트의 원소를 반환하였다. 만약, 풍향이 45.0(m/s) 라면 WS_TAG 값은 45.0이 되도록 설정하였다. 또한, find_WS 함수를 활용해 풍향별 화재 발생 건수를 측정하였다. WS_TAG 컬럼에 find_WS 함수에서 반환된 값을 저장하고, ws_count 리스트를 출력해 풍향별 태그에 해당하는 화재 발생 건수를 출력하였다. 마지막으로, X축은 풍향(m/s), Y축은 화재 발생 건수인 막대그래프를 그려 출력하였다. Fig. 5는 풍향별 화재 발생 현황을 나타낸 것이다. 풍향이 225.0 m/s 인 경우에 화재 발생 가능성이 가장 높다는 것을 알 수 있다.

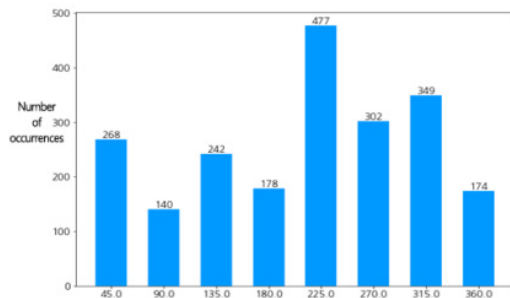


Fig. 5. Current status of fires by wind direction

풍속별 화재 발생 현황 막대그래프 시각화에서는 단위 시간 풍속(TIME_UNIT_WD) 컬럼으로 그룹화하고, fdf 객체에 풍속별 화재 발생 건수 구분을 위한 태그(WD_TAG) 컬럼을 추가하였다. 막대그래프 X축에 풍속을 활용하기 위해 WD_TAG 컬럼을 추가하고,

fdf 객체에 저장된 내용을 출력해 추가된 WD_TAG 컬럼을 확인하였다. 풍속을 표현하기 위한 리스트 생성과 함수(find_WD)를 정의하였다. wd_label 리스트는 막대그래프의 X축에 해당하는 레이블이며, wd_count 리스트는 groups 리스트의 각 원소가 몇 개인지 세는 리스트이다. groups 리스트는 임의로 각각의 풍속을 구분하는 태그이고, find_WD 함수는 value 값으로 행의 TIME_UNIT_WD 값을 받은 뒤, 풍속을 구분하는 태그에 해당하면 wd_count 원소에 1을 더하고 groups 리스트의 원소를 반환하였다. 만약, 풍속이 2.0 이라면 WD_TAG 값은 2.0이 되도록 설정하였다. 또한, find_WD 함수를 활용해 풍속별 화재 발생 건수를 측정하였다. WD_TAG 컬럼에 find_WD 함수에서 반환된 값을 저장하고, wd_count 리스트를 출력해 풍속별 태그에 해당하는 화재 발생 건수를 출력하였다. Fig. 6은 풍속별 화재 발생 현황을 나타낸 것이다. 풍속이 2.0인 경우에 화재 발생 건수가 매우 높은 것을 알 수 있다.

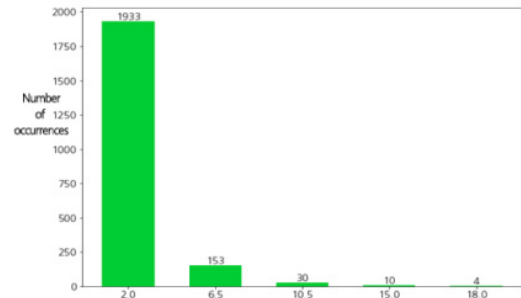


Fig. 6. Current status of fires by wind speed

습도별 화재 발생 현황 막대그래프에서는 단위 시간 습도(TIME_UNIT_HUMIDITY) 컬럼으로 그룹화하고, fdf 객체에 습도별 화재 발생 건수 구분을 위한 태그(HUM_TAG) 컬럼을 추가하였다. 막대그래프 X축에 습도의 범위를 활용하기 위해 HUM_TAG 컬럼을 추가하고, fdf 객체에 저장된 내용을 출력해 추가된 HUM_TAG 컬럼을 확인하였다. 또한, 습도 범위를 표현하기 위한 리스트 생성과 함수(find_HUM)를 정의하였다. hum_label 리스트는 막대그래프의 X축에 해당하는 레이블이며, hum_count 리스트는 groups 리

스트의 각 원소가 몇 개인지 세는 리스트이다. groups 리스트는 임의로 설정한 습도의 범위를 구분하는 태그이며, find_HUM 함수는 value 값으로 행의 TIME_UNIT_HUMIDITY 값을 받은 뒤, 임의로 설정한 습도 범위에 해당하면 hum_count 원소에 1을 더하고 groups 리스트의 원소를 반환하였다. 만약, 20% 라면 11% 이상 20% 이하의 범위에 속하기 때문에 HUM_TAG 값은 20이 되도록 설정하였다. 또한, find_HUM 함수를 활용해 습도별 화재 발생 건수를 측정하였다. HUM_TAG 컬럼에 find_HUM 함수에서 반환된 값을 저장하고, hum_count 리스트를 출력해 습도별 태그에 해당하는 화재 발생 건수를 출력하였다. Fig. 7은 습도별 화재 발생 현황을 나타낸 것이다. 습도가 21% 이상일 때 화재 발생 건수가 많이 발생한 것을 알 수 있으며, 건조할 때보다 화재 발생이 많은 것으로 분석되었다.

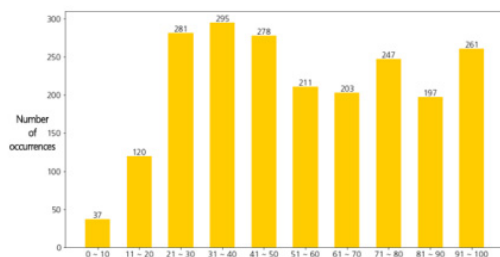


Fig. 7. Fires by humidity

마지막으로, 기상정보에 영향을 받는 화재 위치를 표시하기 위해, 화재 발생 건수가 많았던 기상정보의 조건을 모두 만족하는 화재의 발생 위치를 표시하였다. 기온이 11°C 이상 20°C 이하일 때, 풍향이 225.0m/s 일 때, 풍속이 2.0 일 때, 습도가 31% 이상 40% 일 때 화재 발생 건수가 가장 높았다. 위의 기상정보를 만족하는 화재의 위치를 지오코더를 활용해 표현하였다. Fig. 8은 기상정보의 영향을 받은 화재 발생 위치를 나타낸 것이다.

가장 많은 화재가 발생한 기온, 풍향, 풍속, 습도의 조건을 모두 만족하는 화재를 대상으로 구글 지도맵에 마커로 위치를 표시하였다. 기온이 11°C 이상 20°C 이하, 풍향이 225.0m/s, 풍속이 2.0, 습도가 31% 이상

40% 이하일 때 각 조건별 화재 발생 건수가 높게 도출되었으며, 조건을 모두 만족하는 화재의 위치를 지오코더로 활용하여 표기하였다. 표기된 마커를 클릭하면 해당 화재가 발생한 계절을 출력하도록 설정하였다.



Fig. 8. Marking the location of the fire according to the influence of weather information

4.2 데이터 예측

기상정보에 따른 화재 발생 가능성 예측에서는 데이터 분석에 필요하지 않은 컬럼을 제거하고, 기온별 화재 발생의 상대 빈도를 계산해 기온 리스트에 추가하였다. tmp_arr 리스트는 TMP_TAG 컬럼의 값을 저장하는 리스트이고, tmp_freq 리스트는 tmp_count 리스트(기온 범위에 해당하는 화재 발생 건수)의 각 원소가 상대적으로 어느 정도의 빈도를 차지하는지 계산해서 저장하는 리스트이다. tmp_fire_freq 리스트는 TMP_TAG 컬럼의 값을 상대 빈도로 변환해 저장하는 리스트로 tmp_fire_freq 리스트의 모든 원소를 더한 값은 10이다. 만약, 기온이 8°C 라면 0°C 이상 10°C 이하의 범위에 속해 TMP_TAG가 10이 되고, tmp_freq 리스트에 따르면 빈도는 3.051643192488263이 된다.

또한, 풍향별 화재 발생의 상대 빈도를 계산해 풍향 리스트에 추가하였다. ws_arr 리스트는 WS_TAG 컬럼의 값을 저장하는 리스트이고, ws_freq 리스트는 ws_count 리스트(풍향 범위에 해당하는 화재 발생 건수)의 각 원소가 상대적으로 어느 정도의 빈도를 차지하는지 계산해서 저장하는 리스트이다. ws_fire_freq 리스트는 WS_TAG 컬럼의 값을 상대 빈도로 변환해 저장하는 리스트로 ws_fire_freq 리스트의 모든 원소를 더한 값은 10이다. 만약, 풍향이 45.0m/s 라면 WS_TAG가 45.0이 되고, ws_freq 리스트에 따르면 빈

도는 1.2582159624413145이 된다.

마찬가지로, 풍속별 화재 발생의 상대 빈도를 계산해 풍속 리스트에 추가하였다. wd_arr 리스트는 WD_TAG 컬럼의 값을 저장하는 리스트이고, wd_freq 리스트는 wd_count 리스트(풍속 범위에 해당하는 화재 발생 건수)의 각 원소가 상대적으로 어느 정도의 빈도를 차지하는지 계산해서 저장하는 리스트이다. wd_fire_freq 리스트는 WD_TAG 컬럼의 값을 상대 빈도로 변환해 저장하는 리스트로 wd_fire_freq 리스트의 모든 원소를 더한 값은 10이다. 만약, 풍속이 2.0 이라면 WD_TAG가 2.0이 되고, wd_freq 리스트에 따르면 빈도는 9.075117370892018이 된다.

또한, 습도별 화재 발생의 상대 빈도를 계산해 습도 리스트에 추가하였다. hum_arr 리스트는 HUM_TAG 컬럼의 값을 저장하는 리스트이고, hum_freq 리스트는 hum_count 리스트(습도 범위에 해당하는 화재 발생 건수)의 각 원소가 상대적으로 어느 정도의 빈도를 차지하는지 계산해서 저장하는 리스트이다. hum_fire_freq 리스트는 HUM_TAG 컬럼의 값을 상대 빈도로 변환해 저장하는 리스트로 hum_fire_freq 리스트의 모든 원소를 더한 값은 10이다. 만약, 습도가 20% 라면 11% 이상 20% 이하의 범위에 속해 HUM_TAG가 20이 되고, hum_freq 리스트에 따르면 빈도는 0.5633802816901409 이 된다.

각 기상정보의 상대 빈도를 더하여 FIRE_TAG 컬럼에 저장하였다. 상대 빈도가 저장되어 있는 리스트인 tmp_fire_freq, ws_fire_freq, wd_fire_freq, hum_fire_freq의 각 원소들을 tmp_fire_freq 리스트에 더하여 저장하고, FIRE_TAG 컬럼을 추가한 뒤 tmp_fire_freq 리스트 값을 저장하였다.

최종적으로, 기상 정보에 따른 화재 발생 가능성 예측 모델을 설계하기 위해 선형 회귀 분석 모델을 구축하였다. 우선적으로, 데이터 예측에 필요하지 않은 컬럼을 제거하고, 사이킷런을 사용하여 머신러닝 선형 회귀 분석 모델을 구축하였다. FIRE_TAG 컬럼은 기상정보의 상대 빈도를 더한 값이 저장되어 있다. 기상정보를 입력하면 FIRE_TAG의 값을 출력하는 것이 프로젝트의 목표이므로, FIRE_TAG 피처를 회귀식의 종속 변수 Y로 설정하고 나머지를 독립 변수 X로 설

정하였다. 데이터를 7:3 비율로 분할하여 학습 데이터와 평가 데이터로 설정하고, 선형 회귀 분석 모델 객체인 lr을 생성한 뒤 학습 데이터 X_train과 Y_train으로 학습을 수행하였으며, 평가 데이터 X_test로 예측을 수행해 예측값 Y_predict를 도출하였다.

또한, 평가 지표를 통해 선형 회귀 분석 모델 평가와 회귀 계수를 확인하였다. squared_error()를 이용해 평가 지표 MSE를 구하고 구한 값의 제곱근을 계산하여 평가지표 RMSE를 구하고, r2_score()를 이용해 평가지표 R²를 구하고, 선형 회귀의 Y절편과 각 피처의 회귀 계수를 확인하였다. 회귀 모델에서 구한 회귀 계수 값과 피처 이름을 묶어 Series 자료형으로 만들고, 회귀 계수 값을 기준으로 내림차순 정렬해 회귀 계수 값이 큰 항목을 확인하였다.

마지막으로, 회귀 분석 결과를 산점도와 선형 회귀 그래프로 출력하였다. subplots()를 사용해 독립변수인 4개 피처와 종속 변수인 FIRE_TAG와의 회귀 관계를 보여주는 4개 그래프를 2행 2열 구조로 표현하였다. Fig. 9는 산점도와 선형 회귀 그래프를 활용하여 기상 정보에 따른 화재 발생 가능성을 예측한 결과이다.

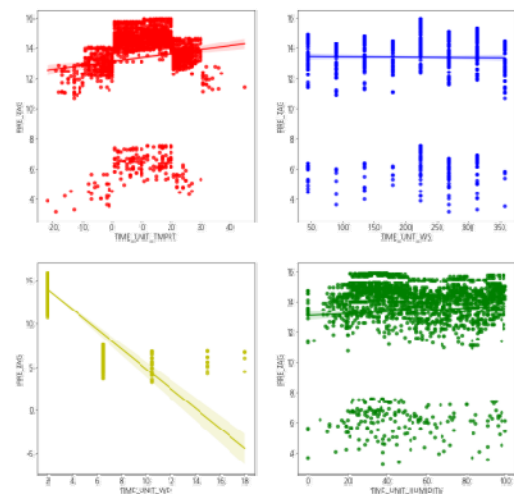


Fig. 9. Scatterplot and linear regression graph results

이를 토대로, 기상정보를 입력받아 화재 발생 가능성을 예측하였다. 4개의 독립 변수(기온, 풍향, 풍속, 습도) 값을 입력하면 변수에 저장한 뒤 회귀 모델에 적

용하여 최종 예측 결과를 도출하였다. Fig. 10은 최종 예측 결과 화면을 나타낸 것이다.

```
print("기상정보를 입력해 화재를 예측하세요.")
tmp = int(input("기온: "))
ws = float(input("풍속: "))
wd = float(input("풍향: "))
hum = int(input("습도: "))

기상정보를 입력해 화재를 예측하세요.
기온: 25
풍속: 225.0
풍향: 2.0
습도: 34

fire_predict = lr.predict([[tmp, ws, wd, hum]])

print("입력한 기상정보에 따른 화재 발생 가능성은 %.2f입니다."%fire_predict)
입력한 기상정보에 따른 화재 발생 가능성은 14.40입니다.
```

Fig. 10. Prediction result of fire according to input weather information

V. 결론

화재는 언제 어디서든지 발생할 수 있는 것으로 막대한 피해를 불러일으킬 수 있다. 화재는 다양한 원인에 의해 발생하며, 이로 인해 막대한 물질, 인적 피해를 초래하고 있다. 따라서, 사전에 화재 발생 가능성을 예측하고 피해를 최소화하기 위한 노력이 필요하다. 화재 발생과 관련하여 다양한 요소 중 기상 정보에 따라 화재 발생 및 규모에 영향을 받을 수 있다. 특히, 우리나라의 경우, 고온다습한 여름에는 화재 발생 건수가 감소하고, 저온저습한 겨울철에는 화재 발생 건수가 증가한다. 일반적으로 기온이 낮고 습도가 낮은 건조한 날씨에 주로 화재가 발생한다.

따라서, 본 논문에서는 화재 사고 원인 분석, 계절별 화재 발생률 등을 분석하고 기온, 풍향, 풍속, 습도와 같은 기상 정보와 화재 발생 건수와의 연관성을 분석하였다. 이를 통해, 기상 정보 입력으로 화재 발생 가능성을 예측하는 모델을 설계하였다.

기상 정보와 화재 발생과의 상관성 분석 결과, 기온이 0°C 이상 20°C 이하인 경우, 화재 발생 가능성(FIRE_TAG)이 높았으며, 풍향이 225.0m/s 인 경우, 화재 발생 가능성(FIRE_TAG)이 높게 나타났다. 또한, 풍속이 2.0 인 경우, 화재 발생 가능성(FIRE_TAG)이 높았으며, 습도가 20% 이상 50% 이하인 경우, 화재 발생 가능성(FIRE_TAG)이 높은 것으로 나타났다.

결국, 기상 정보와 화재 발생 간에 관련이 있음을 알 수 있다. 이를 기반으로, 기상정보를 입력받아 화재 발생 가능성을 예측하였다. 4개의 독립 변수(기온, 풍향, 풍속, 습도) 값을 입력하면 변수에 저장한 뒤 회귀 모델에 적용하여 최종 예측 결과를 도출하는 모델을 설계하였다.

REFERENCES

- [1] D. C. Choi, H. S. Hwang, M. H. Ko & S. Y. Lee. (2020). A study on fire and evacuation simulation analysis for use of disaster vulnerable personal evacuation device. *Journal of the Society of Disaster Information*, 16(4), 824-831. DOI : 10.15683/KOSDI.2020.12.31.824
- [2] J. G. Kim, M. S. Jeong & J. W. Jung. (2022). A study on the external evacuation system for large-scale fire of multi-use facilities. *Journal of the Society of Disaster Information*, 18(1), 129-145. DOI : 10.15683/kosdi.2022.3.31.129
- [3] F. Ahmad & L. Goparaju. (2019). Forest fire trend and influence of climate variability in India: A geospatial analysis at national and local scale. *Ekolandoacute;gia (Bratislava)*, 38(1), 49-68. DOI : 10.2478/eko-2019-0005
- [4] J. G. Kim, C. I. Park, J. W. Jung & S. G. Kim. (2021). A study on the risks factors of fire occurrence and expansion for traditional markets. *Journal of the Society of Disaster Information*, 17(1), 60-67. DOI : 10.15683/KOSDI.2021.3.31.060
- [5] S. Y. Hong, S. H. Cho, M. S. Kim & I. Moon. (2019). Fire prediction based on weather and fire data using artificial neural network. *Journal of the Korean Society of Hazard Mitigation*, 19(7), 275-281. DOI : 10.9798/KOSHAM.2019.19.7.275
- [6] B. H. Kim. (2021). Design of reliable disaster

- recovery system through integrated server redundancy. *Turkish Journal of Computer and Mathematics Education*, 12(6), 674-679.
- [7] L. M. Spineli & N. Pandis. (2020). Exploring heterogeneity in meta-analysis: Meta-regression analysis. *American Journal of Orthodontics and Dentofacial Orthopedics*, 158(4), 623-625. DOI : 10.1016/j.ajodo.2020.07.002
- [8] J. Wang & X. Lin. (2020). A Bayesian approach for semiparametric regression analysis of panel count data. *Lifetime Data Analysis*, 26(2), 402-420. DOI : 10.1007/s10985-019-09471-3
- [9] N. S. Kwark & C. H. Lee. (2021). Asymmetric effects of financial conditions on GDP growth in Korea: A quantile regression analysis. *Economic Modelling*, 94, 351-369. DOI : 10.1016/j.econmod.2020.10.014
- [10] B. H. Kim & J. H. Cho. (2020). A study on modular smart plant factory using morphological image processing. *Electronics*, 9(10), 1661. DOI : 10.3390/electronics9101661
- [11] B. H. Kim. (2022). A study on the prediction of KRW/USD exchange rate against the number of COVID-19 confirmed cases using LSTM model. *Journal of Next-generation Convergence Technology Association*, 6(4), 593-598. DOI : 10.33097/JNCTA.2022.06.04.593
- [12] D. Shai. (2006). Income, housing and fire injuries: A census tract analysis. *Public Health Reports*, 121(2), 149-154. DOI : 10.1177/003335490612100208
- [13] W. Stanley & S. Gilbert. (2018). *Community risk ratings for the FireCARES system*. US Department of Commerce National Institute of Standards and Technology.
- [14] M. Taylor, D. Appleton, G. Oakford & J. Fielding. (2020). Population trends and fire prevention in merseyside UK. *Fire Technology* 2020.
- [15] S. Volkan, K. Omer & G. Merih. (2020). A Bayesian network model for prediction and analysis of possible forest fire causes. *Forest Ecology and Management*, 457, 117723.
- [16] K. R. Singh, K. P. Neethu, K. Madhurekaa, A. Harita & P. Mohan. (2021). Parallel SVM model for forest fire prediction. *Soft Computing Letters*, 3, 100014.
- [17] K. S. Ko, J. K. Yang, D. H. Hwang, H. S. Ko, C. O. Ga & J. P. Cho. (2020). Building fire prediction model study using AI. *The Journal of Korean Institute of Communications and Information Sciences*, 45(7), 1210-1218. DOI : 10.7840/kics.2020.45.7.1210
- [18] K. S. Ko. (2018). Electrical fire prediction model study using machine learning. *Journal of KIIECT*, 11(6), 703-710. DOI : 10.17661/jkiiect.2018.11.6.703
- [19] H. Jung & J. Kim. (2017). A machine learning approach for mechanical motor fault diagnosis. *Society of Korea Indus. and Sys. Eng.*, 40(1), 57-64. DOI : 10.11627/jkise.2017.40.1.057

저 자 약 령

김봉현(Kim, Bong Hyun)

[회원]



- 2009년 2월 : 한밭대학교 컴퓨터 공학과(공학박사)
- 2012년~2015년 : 경남대학교 컴퓨터공학과 교수
- 2020년~현재 : 서원대학교 소프트웨어학부 컴퓨터공학과 교수
- E-Mail : bhkim@seowon.ac.kr

<관심분야>

Big data, IoT, AI Service, ICT convergence