

기상현상과 화재 발생에 대한 상관분석

참가번호

240366

팀 명

ABLE

1. 분석 배경 및 목표

기후 변화로 인해 기상 조건이 화재 발생에 미치는 영향이 커지고 있다. 높은 온도, 강한 바람, 낮은 습도 등은 화재 발생 가능성을 높일 수 있는 주요 요인으로 추정된다. 이러한 기상 조건이 실제로 화재 발생에 어떤 영향을 미치는지에 대한 정확한 분석이 필요하다. 이번 분석의 목표는 다음과 같다.

- 기상 데이터 수집 및 분석: 제공받은 온도, 습도, 강수량, 바람 세기 등의 기상 데이터를 분석하여 화재 발생에 미치는 영향을 파악한다.
- 화재 데이터 수집 및 분석: 제공받은 지역별, 시간대별 화재 발생 데이터를 분석하고, 통계적 패턴을 도출한다.
- 기상현상과 화재 발생 간의 상관관계 분석: 기상 데이터와 화재 데이터 및 특보 데이터를 결합하여 상관분석을 실시하고, 특정 기상 조건 및 특보가 화재 발생에 미치는 영향을 확인한다.

이 분석을 통해 날씨와 화재 발생 간의 상관관계를 명확히 파악하는 것이 목표이다.

2. 분석 데이터 정의

본 공모안에서는 기상, 화재발생, 기상특보, 데이터를 사용하였다. 화재발생 데이터의 Fire_type_2(화재 유형 소분류)에서 결측치가 발견되었고, 기상 데이터에서는 -99.0의 이상값을 확인하였다. 처음에 세 데이터를 하나로 병합했으나 의미 있는 상관분석 결과를 얻지 못해 병합 순서에 따라 (화재+기상), (기상+화재), (화재+기상)으로 나누어 상관분석을 진행하였다. 기상 데이터의 -99.0 결측값은 평균치 또는 선형보간으로 처리해 보았으나 실제 데이터를 사용하는 것이 더 낫다고 판단하여, 실제 지역별 기상청 데이터를 수집하여 결측값을 직접 처리하였다.

[표 1] 화재발생 데이터 소개

화재발생 데이터																
컬럼명	NUM	Year	TM	district_1	district_2	Fire_ty pe_1	Fire_ty pe_2	Igni tion_fa cto_r_C ate_gor y_1	Igni tion_fa cto_r_C ate_gor y_2	casualties	De ad	Inju ry	Pro per_ty_ da ma ge	locatio n_C ate gor y_1	lacatio n_C ate gor y_2	locatio n_C ate gor y_3
내용	연 번	연 도	화 재 발 생 년 월 일	시 도	시 군 구	화 재 유 형 대 분 류	화 재 유 형 소 분 류	발 화 요 인 대 분 류	발 화 요 인 소 분 류	인 명 피 해 소 계	사 망	부 상	재 산 피 해 소 계	장 소 대 분 류	장 소 중 분 류	장 소 소 분 류

[표 2] 기상 데이터 소개

기상 데이터															
컬럼명	TM	dist rict _1	dist rict _2	ST N	ta_ ma x	ta_ min	ta_ min _m ax	rn_ day	ws_ ma x	ws_ ins_ ma x	ws_ me an	ws_ min	hm_ ma x	hm_ me an	hm_mi n
내용	날 짜	시 도 명	시 군 구 명	AW S 지 점 코 드	일 최 대 기 온	일 최 저 기 온	일 교 차	일 강 수 량	일 최 대 풍 속	일 최 대 순 간 풍 속	일 평 균 풍 속	일 최 저 풍 속	일 최 대 상 대 습 도	일 평 균 상 대 습 도	일 최 저 상 대 습 도

[표 3] 기상 특보 데이터 소개

기상 특보 데이터									
컬럼명	fire_wea ther_sp ecial.tm _fc	fire_wea ther_sp ecial.tm _ef	fire_wea ther_sp ecial.dis trict_1	fire_wea ther_sp ecial.dis trict_2	fire_wea ther_sp ecial.stn	fire_wea ther_sp ecial.re g_id	fire_wea ther_sp ecial.wr n	fire_wea ther_sp ecial.lv l	fire_wea ther_sp ecial.cm d
내용	특보시 작	특보종 료	대 지역	소 지역	AWS 지점 코드	특보구 역 코드	특보 종류	특보 단계	특보 명령

3. 데이터 전처리 및 변수 추가

● 기존 기상 데이터 결측값 처리

- 기존 기상 데이터의 결측치를 실제 지역별 기상청 데이터를 활용하여 대체하였다. 대회에서 제공받은 기상 데이터의 '시도' 컬럼에는 '강원특별자치도'의 값이 있었지만, 실제 강원 지역의 기상청 데이터는 '강원 영서'와 '강원 영동'으로 나뉘어 있었다. 따라서 '시군구명' 컬럼을 기반으로 '강원특별자치도' 데이터를 '강원 영서'와 '강원 영동'으로 변환하였다. 또한, 세종특별자치시의 데이터는 2019년 4월 이후만 존재하였으므로, 이전 데이터는 지역적으로 가까운 충남의 기상 데이터로 대체하였다.

● 화재 발생 + 기상 데이터

새로 생성된 데이터에는 지역명이 '강원영동', '강원영서'로 분리되어 있었지만, 화재발생데이터와 병합하기 위해 '강원특별자치도'라는 새로운 값으로 대체하였다. '시도명', '시군구명'을 묶어 '지역'이라는 새로운 컬럼을 생성하였고, '화재발생시간'을 '일시', '월', '시간'별로 분석하기 위해 3가지 컬럼으로 생성하였다.

화재발생 데이터에서 대체할 수 있는 '화재발생시간', '연번'을 삭제하고, 결측치가 존재하는 '화재유형 소분류' 컬럼을 '화재유형 대분류' 컬럼의 값과 연관성이 높아 삭제시켰다. 기상 데이터도 마찬가지로 '시군구명', 'AWS 지점 코드'를 삭제하였다.

화재발생날짜를 기준으로 해당 날의 날씨와 병합하기 위해 '지역', '일시'별로 left 병합을 실행하였다. 건조여부가 화재 발생에 중요한 관계가 있을 것으로 추정되어 찾아본 결과 강수량이 0이고 실험습도가 30% 미만인 경우에 건조하다고 판단하는 것이 기상청의 건조 특보 기준과 대체로 일치하여 건조여부 컬럼을 생성하였다.

계절마다 화재 여부를 파악하기 위해 '계절' 컬럼을 생성 후 '봄(3~5)', '여름(6~8)', '가을(9~11)', '겨울(12~2)'로 구별하였다.

시간별 화재 여부를 파악하기 위해 '시간대' 컬럼을 생성 후 '오전(0~12)', '오후(12~18)', '밤(18~24)', '새벽(0~6)'이라는 컬럼을 생성하였다.

● 특보 + 화재발생 데이터

특보 데이터의 '지역 ID' 컬럼과 추가로 수집한 특보구역명 데이터의 '코드' 컬럼을 left 병합하여 merged_df 데이터를 생성하였다. 이후 '대지역' 컬럼과 '소지역' 컬럼을 합쳐서 '위치' 컬럼을 만들었다. 결측값이 존재하는 '코드' 컬럼과 '특보구역명' 컬럼에 대해서 네이버 지도를 활용하여 대략적인 위치를 파악하고, 이를 기반으로 결측값을 업데이트하였다. 또한 '날짜' 컬럼에서 시간 데이터를 제외하여 '년-월-일' 컬럼을 만들었다.

확보한 화재 데이터의 '대지역' 컬럼과 '소지역' 컬럼을 합쳐서 '위치' 컬럼을 만들었고, '날짜' 컬럼에서 시간 데이터를 제외하여 '년-월-일' 컬럼을 만들었다. 이후 화재 데이터와 merged_df를 '년-월-일' 컬럼과 '위치' 컬럼으로 left 병합하여 merged 데이터를 생성하였다. 특보가 진행 중인 시간에 발생한 화재만 확인하기 위해 '특보시작'과 '특보종료' 사이에 발생한 데이터로 필터링하여 filtered 데이터를 생성하였다. 이후 상관분석에 필요한 컬럼만 남기고 나머지 컬럼을 삭제하였다.

분석을 위해 merged 데이터에 '화재여부' 컬럼을 생성하고, filtered 데이터에 대해서 화재 여부를 1로 설정하였으며, 그 외의 행은 0으로 채웠다. 경보 데이터에 따른 화재 종류에 대한 상관관계를 분석하기 위해 특정 조건에 해당하는 경우 '대형화재', '중요화재', '특수화재' 컬럼을 각각 만들었다. 특정 조건에 해당하지 않는 경우에는 '일반화재'에 해당하는 컬럼을 생성하여 값을 부여하였다.

● 기상 + 화재발생 데이터

앞서 진행한 (화재발생 + 기상) 데이터와 마찬가지로 동일한 데이터 전처리 및 변수 추가를 진행하였다. 기온이 높으면 화재 발생 가능성이 높아지고 풍속이 강하면 화재의 규모가 커지는 특성을 갖고 있기 때문에 두 요인이 함께 작용한다면 화재 위험성을 판단할 수 있다고 생각하여 '기온'과 '습도' 컬럼의 값을 곱하여 '화재예측지표' 컬럼을 생성하였다.

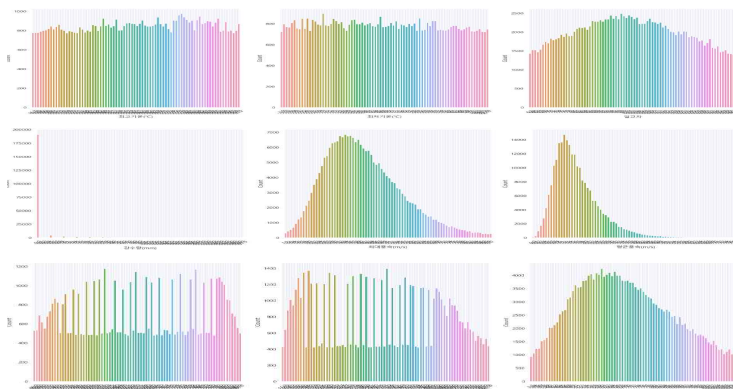
4. 상관분석 및 시각화

● 화재 + 기상



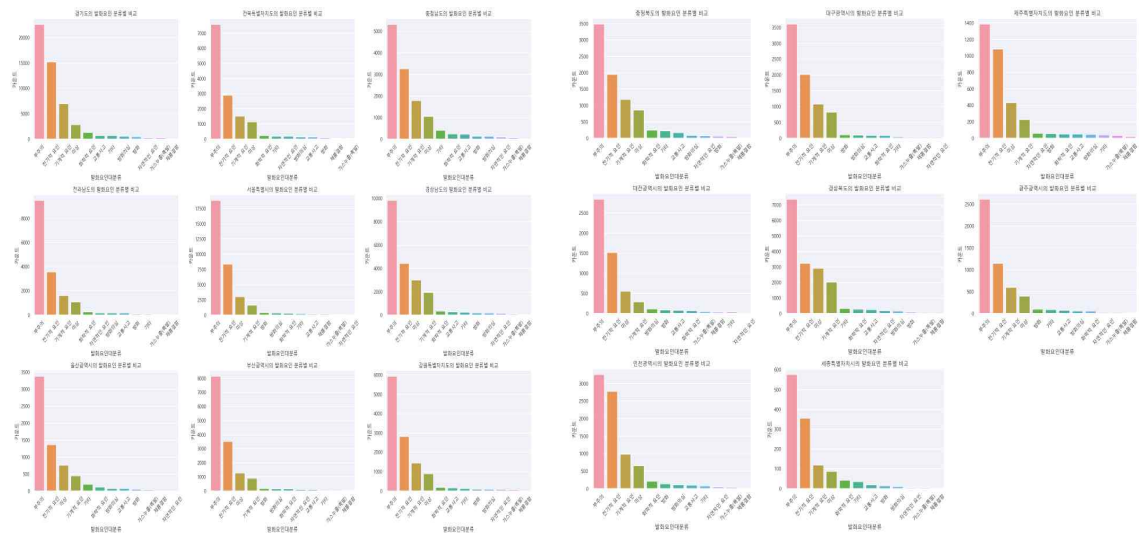
[그림 1] 화재가 발생하는 것은 건조 여부와 관계가 있다고 생각하여 '건조 여부' 컬럼을 생성 후 지역별 시각화를 해본 결과 건조 여부는 화재 발생과 연관성이 없는 것으로 추측하였다.

[그림 1] 화재발생일 지역별 건조 그래프



[그림 2] 기상별 화재 비교

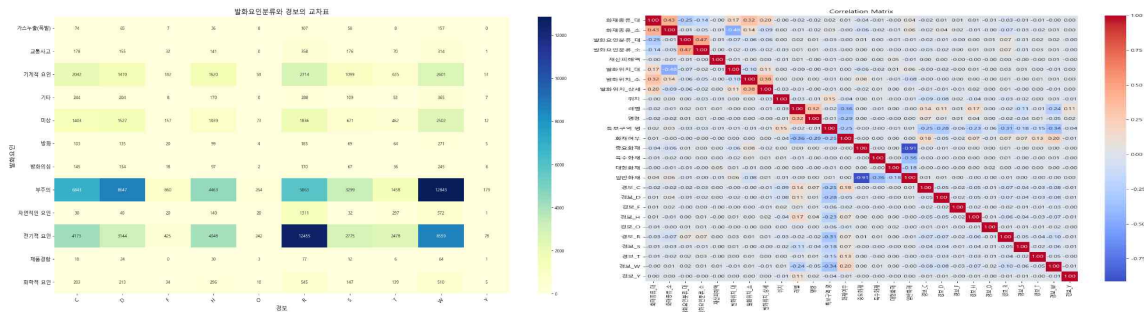
[그림 2] 기상별로 화재빈도 수를 파악해 본 결과 최고기온, 최저기온에 비해 일교차, 강수량(mm), 풍속(m/s), 습도(%/rh)가 특정 값 사이에서 유의미한 상관성이 있음을 확인하였다. '습도' 관련 컬럼에서는 결측값 2988개를 없앴기 때문에 비어있는 데이터 모습이다.



[그림 3] 지역별 화재발생요인 그래프

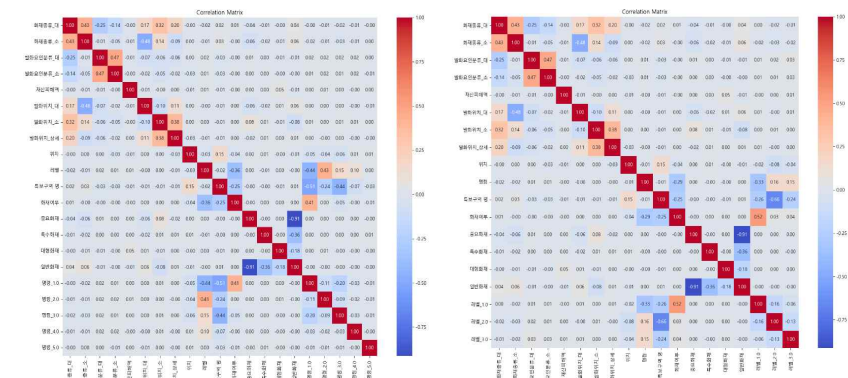
[그림 3] 지역별 화재발생 요인 확인 결과 "부주의" 혹은 "전기적 요인"으로 인한 화재가 공통적으로 화재발생에 높은 비중을 차지했고, 그에 비해 자연적인 요인은 미미한 값을 보여줌을 확인하였다.

● 특보 + 화재



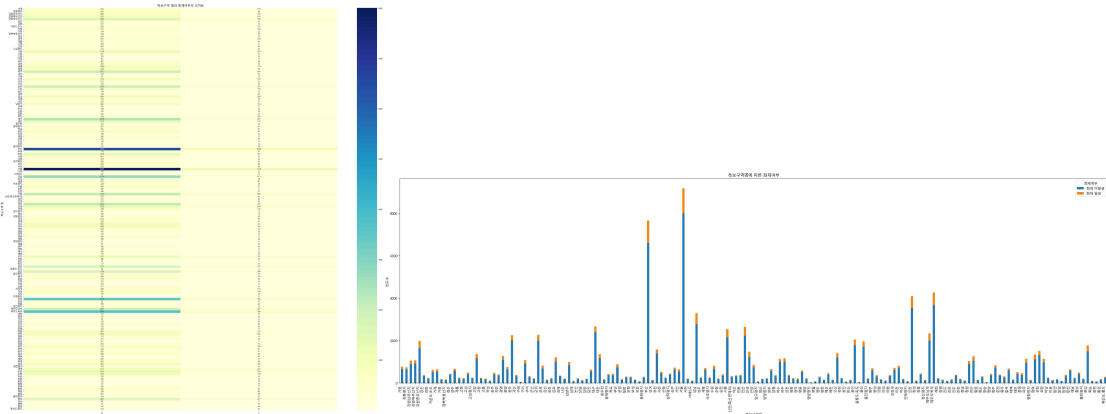
[그림 4] 기상 특보 데이터와 화재 데이터 교차표 및 경보 컬럼과 화재 여부 상관분석

[그림 4] '경보' 컬럼을 원-핫 인코딩하여 merged 데이터와 상관 분석한 결과 W(강풍) 0.20, C(한파) 0.18, T(태풍) 0.13으로 화재 여부와 상관관계가 있음을 확인하였고, '발화요인 분류_대'와 '경보' 컬럼의 교차표를 활용해 D(건조)와 W(강풍) 특보일 때 부주의로 인한 화재 발생, R(호우)와 W(강풍) 특보일 때 전기적 요인의 화재가 자주 발생한 것을 각각 확인하였다.



[그림 5] 명령 컬럼, 레벨 컬럼과 화재 여부 상관분석

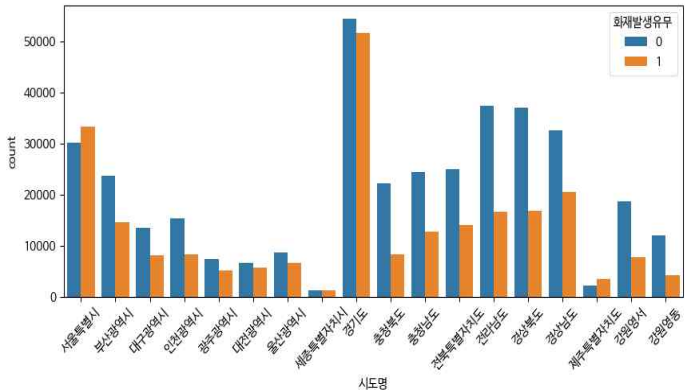
[그림 5] '명령' 컬럼을 원-핫 인코딩하여 merged 데이터와 상관 분석한 결과 명령 1(예비) 0.41로 화재 여부와 상관관계가 있음을 확인하였다. '레벨' 컬럼을 원-핫 인코딩하여 merged 데이터와 상관 분석한 결과 레벨 1(발표) 0.52로 화재 여부와 상관관계가 있음을 확인하였다.



[그림 6] 특보구역명별 화재여부 교차표 및 막대그래프 시각화

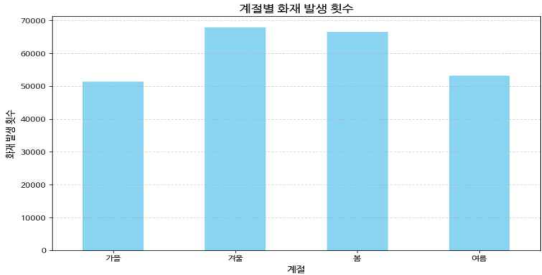
[그림 6] '특보구역명'을 그룹화하여 '화재여부' 컬럼과 이변량 분석을 교차표와 막대그래프를 활용하여 부산과 서울에서 구역별 화재 빈도수가 가장 높음을 확인하였다.

● 기상 + 화재

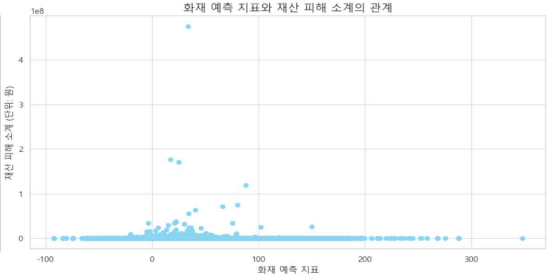


[그림 7] 시도별 화재 발생 그래프를 확인해 본 결과 경기도 지역의 화재 발생이 가장 높은 것을 확인할 수 있었고, 전체적으로 화재가 발생한 경우보다 발생하지 않은 경우가 더 많은데 서울특별시는 화재가 발생한 경우가 더 많음을 확인하였다.

[그림 7] 시도별 화재발생 유무



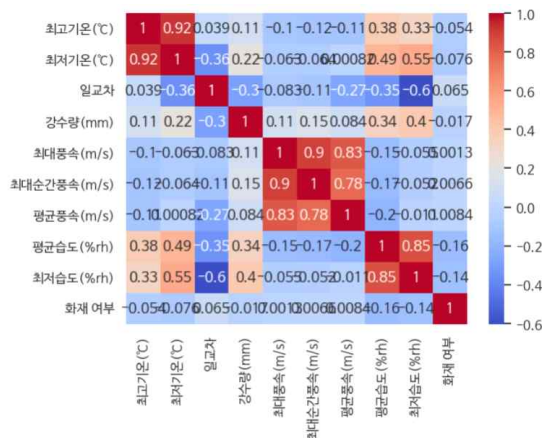
[그림 8] 계절별 화재 발생 횟수



[그림 9] 화재 예측 지표와 재산 피해 관계

[그림 8] 날씨가 건조한 계절에 화재 발생률이 높을 것이라는 추측으로 계절별 화재 발생을 그래프로 그려본 결과 봄과 겨울이 많은 화재가 발생했고, 가을과 여름이 비교적 화재가 적게 발생하였다.

[그림 9] 기온과 풍속을 곁한 화재예측지표와 재산피해소계의 관계를 파악하기 위해 산점도로 시각화한 결과 두 컬럼이 뚜렷한 경향성을 보이지 않아 두 변수 간에는 상관관계가 부재하고 직접적인 관련성이 없다고 판단하였다.



[그림 10] 기상 정보와 화재발생 유무 상관관계 히트맵

[그림 10]에서는 기상정보와 화재발생유무 간의 상관관계수를 파악하기 위해 히트맵 시각화를 수행한 결과를 보여준다. 이를 통해 기상상태와 화재발생 간에는 약한 상관관계가 있음을 확인할 수 있었다. 예를 들어, 일교차가 클수록 화재발생이 더 많은 경향을 보이며, 풍속이 강한 날에는 화재가 발생할 가능성이 높아지는 등의 경향을 발견할 수 있었다. 이 결과를 통해 화재발생에 영향을 미치는 기상 조건을 조금 더 자세히 이해할 수 있게 되었고, 이를 토대로 화재 예방 및 대응 정책을 더욱 효과적으로 수립할 수 있을 것으로 기대된다.

5. 모델링 및 예측

0.6809941536429192				
	precision	recall	f1-score	support
0	0.70	0.82	0.76	111611
1	0.63	0.46	0.53	71580
accuracy			0.68	183191
macro avg	0.66	0.64	0.64	183191
weighted avg	0.67	0.68	0.67	183191

[그림 11] 기상에 따른 화재 여부 예측 결과

[그림 11] 화재 여부와 기상정보는 약한 상관관계를 가지고 있지만 약한 상관관계를 가지고 있다고 화재여부를 예측할 수 없는 것인지 확인해 보기 위해 화재 여부 예측 모델을 만들었다. 모델은 랜덤 포레스트 모델을 100개의 결정트리로 학습시켰다. 성능테스트 결과 0.68의 정확도를 보이는 것을 확인할 수 있었다.

6. 활용 방안 및 기대효과

위 분석을 토대로 약한 상관관계를 가진 기상 조건인 일교차, 강수량, 풍속, 습도에 따라 화재 예방과 대비 태세를 강화하고, 기상 예보를 바탕으로 화재 위험성을 평가하여 사전에 소방 인력과 장비를 배치함으로써, 화재 발생률을 낮추고, 신속한 대응으로 피해를 최소화하며, 데이터 기반의 효과적인 화재 예방 및 대응 정책을 수립할 수 있다.

예를 들어, 강풍 예보 시에는 전기적 요인에 의한 화재 예방을 위한 전기 설비 점검을 실시하고, 습도가 낮은 날에는 화재 발생 가능성을 경고하는 메시지를 배포합니다. 이는 보다 효율적이고 효과적인 화재 관리 체계를 구축할 것으로 기대한다.

※ 참고문헌

- 기상청 공공 데이터

[공공데이터개방안내 > 정보공개 > 기상청 \(kma.go.kr\)](http://kma.go.kr)

- 세종시 종관기상관측장비 설치

[세종시에 종관기상관측장비 설치...예비관측 시작 | 연합뉴스 \(yna.co.kr\)](http://yna.co.kr)

- 화재조사 및 보고규정

[화재조사및보고규정 | 국가법령정보센터 | 행정규칙 \(law.go.kr\)](http://law.go.kr)