

AI 연기 감지 모델을 활용한 스마트 화재 예측 시스템

현대건설 기술교육원 Smart 안전 4 조

발표자: 이승연(보건)

팀원: 강혜나, 김민준, 정세림

Github: [MJ0817/Smoke_Detection \(github.com\)](https://github.com/MJ0817/Smoke_Detection)

1. 프로젝트 개요

- 1) IoT 장치에서 수집된 데이터를 기반으로 화재 발생 여부를 예측하는 머신러닝 모델을 개발하는 것을 목표로 한다. 다양한 환경 변수를 활용하여 연기감지기를 통해 조기에 화재를 감지하고, 실시간 경보시스템과 통합하여 화재 피해를 최소화하는데 있다.

2. 개발 목적

- 1) 머신러닝 모델 활용 대상
 - IoT 환경에서 실시간 화재 예측 시스템에 활용될 수 있다. 다양한 센서를 통해 수집된 데이터를 실시간으로 분석하여 화재 발생 가능성을 예측하고 필요시 경보를 울릴 수 있는 시스템에 통합하여 사용할 수 있다.
- 2) 개발 의의
 - 화재를 조기에 감지함으로써 인명과 재산 피해를 줄이는 데 중요한 역할을 할 것으로 예상된다.
 - 공공장소 및 대규모 산업시설 등에서 실시간 모니터링 시스템과 결합하여 안전성 향상에 기여할 수 있다.

3. 배경지식

1) 데이터 관련 사회 문제

- 2024 년 8 월 1 일 청라국제도시 아파트 단지 지하주차장에서 전기차 화재가 발생하였는데 이로 인해 140 대 이상의 차량 손실과 주민 22 명이 병원으로 이송되는 등 재산 및 인명 피해가 발생하였다. 현재 건축법 시행령과 소방시설 설치 및 관리에 관한 법률 시행령에서는 지하주차장에 배연설비와 제연설비를 설치하는 것을 의무사항으로 두고 있지 않다. 이와 같은 상황에서 재발 방지를 위한 실질적인 개선이 필요한 상태이다.

4. 개발 내용

1) 독립변수와 종속변수 설정

- 독립변수(입력 변수): 온도, 습도, TVOC, eCO₂, 센서가 감지한 수소와 에탄올 농도, 공기압, 미세먼지 농도 등 다양한 환경 변수들이 포함됩니다. 이 변수들은 화재 발생 여부에 영향을 줄 수 있는 다양한 요인들을 나타낸다.
- 종속변수(목표 변수): 화재 경보(Fire Alarm). 이 변수는 화재가 발생했는지 여부를 이진(0과 1)으로 나타내며, 예측의 목표가 된다.

2) 데이터 세부정보 및 주요속성

- 온도(Temperature[C]): 주위 환경의 온도를 나타내는 연속형 변수이다. 화재 발생 시 온도의 변화는 중요한 지표가 될 수 있다.
- 습도(Humidity[%]): 공기 중 수분의 양을 나타내는 연속형 변수이다. 습도는 화재 발생 가능성에 영향을 줄 수 있는 중요한 요소이다.
- 총휘발성유기화합물(TVOC[ppb]): 공기 중 휘발성 유기 화합물의 농도를 나타내는 변수로, 화재 시에 급격히 증가할 수 있다.
- 이산화탄소 농도(eCO₂[ppm]): CO₂ 농도는 화재 발생 시 공기 중의 변화로 나타날 수 있다.

- 센서가 감지한 수소 농도(Raw H2): 수소의 농도를 나타내는 변수로, 화재와 관련된 화학 반응에서 변화할 수 있다.
- 센서가 감지한 에탄올 농도(Raw Ethanol): 에탄올의 농도를 나타내는 변수이다. 특정 화재 시나리오에서 농도가 변화할 수 있다.
- 공기압(Pressure[hPa]): 대기압을 나타내는 변수로, 화재 발생 시 대기압의 미세한 변화가 감지될 수 있다.
- 미세먼지 농도(PM1.0, PM2.5): 공기 중 미세먼지의 농도를 나타내는 변수로, 화재 시 발생하는 연기 입자에 따라 값이 변화할 수 있다.
- 화재 경보(Fire Alarm): 화재가 발생했는지를 나타내는 이진형 변수이다(0: 화재 아님, 1: 화재 발생).

	UTC	Temperature	Humidity	TVOC[ppb]	CO2[ppm]	Raw H2	Raw Ethanol	Pressure[hPa]	PM1.0	PM2.5	NC0.5	NC1.0	NC2.5	CNT	Fire Alarm
0	1.65E+09	20	57.36	0	400	12306	18520	939.735	0	0	0	0	0	0	0
1	1.65E+09	20.015	56.67	0	400	12345	18651	939.744	0	0	0	0	0	1	0
2	1.65E+09	20.029	55.96	0	400	12374	18764	939.738	0	0	0	0	0	2	0
3	1.65E+09	20.044	55.28	0	400	12390	18849	939.736	0	0	0	0	0	3	0
4	1.65E+09	20.059	54.69	0	400	12403	18921	939.744	0	0	0	0	0	4	0
5	1.65E+09	20.073	54.12	0	400	12419	18998	939.725	0	0	0	0	0	5	0
6	1.65E+09	20.088	53.61	0	400	12432	19058	939.738	0	0	0	0	0	6	0
7	1.65E+09	20.103	53.2	0	400	12439	19114	939.758	0	0	0	0	0	7	0
8	1.65E+09	20.117	52.81	0	400	12448	19155	939.758	0	0	0	0	0	8	0
9	1.65E+09	20.132	52.46	0	400	12453	19195	939.756	0.9	3.78	0	4.369	2.78	9	0
10	1.65E+09	20.146	52.15	0	400	12454	19230	939.757	0.89	3.71	0	4.289	2.73	10	0
11	1.65E+09	20.161	51.84	0	400	12467	19264	939.754	0.84	3.51	0	4.053	2.58	11	0
12	1.65E+09	20.175	51.62	0	400	12467	19299	939.755	0.81	3.38	0	3.909	2.488	12	0
13	1.65E+09	20.19	51.39	0	400	12469	19317	939.758	0.74	3.11	0	3.588	2.284	13	0
14	1.65E+09	20.204	51.17	0	403	12468	19338	939.742	0.71	2.96	0	3.419	2.176	14	0
15	1.65E+09	20.219	50.99	0	400	12475	19362	939.741	0.64	2.66	0	3.077	1.959	15	0
16	1.65E+09	20.233	50.86	0	400	12480	19382	939.758	0.6	2.52	0	2.908	1.851	16	0

사진 1: 데이터 파일표

2) 데이터 간의 상관관계는 히트맵(Heatmap)을 통해 시각화하여 분석했다. 히트맵은 변수 간의 상관관계를 색상으로 표현하여, 변수들이 얼마나 관련이 있는지를 쉽게 파악할 수 있게 한다.

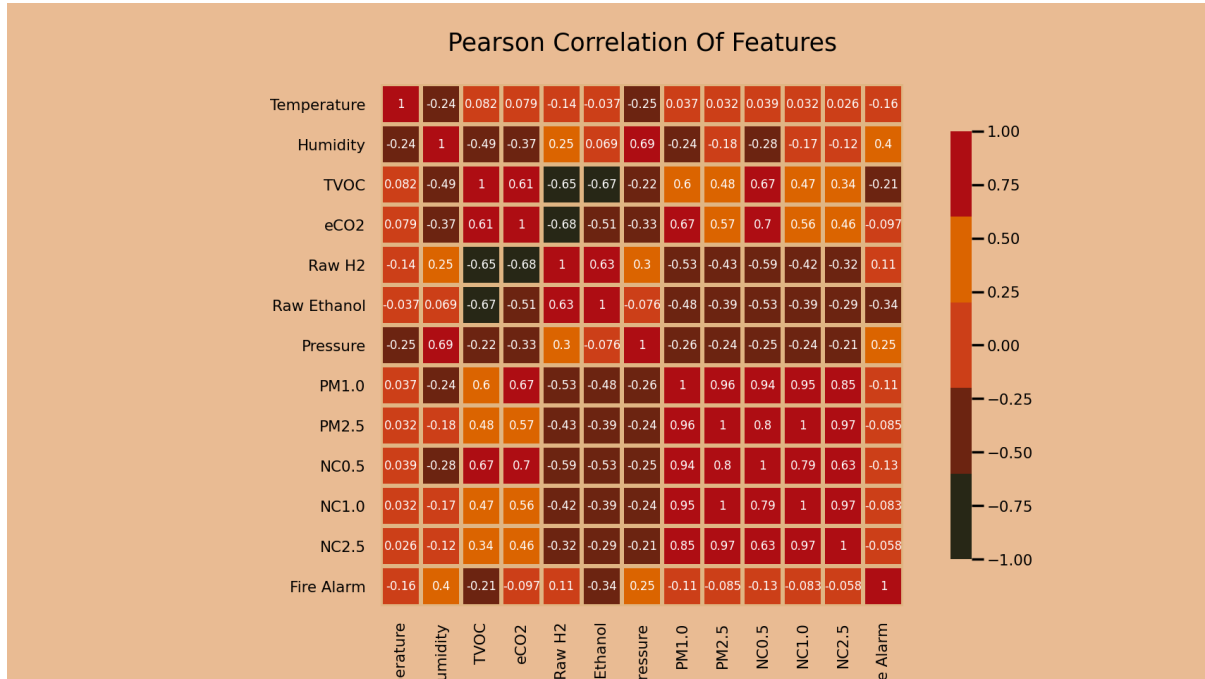


사진2: 데이터 간의 상관관계를 나타내는 히트맵

- 온도와 습도는 반비례 관계를 보여줍니다. 온도가 높아질수록 습도가 낮아지는 경향이 있다.
- TVOC와 eCO2는 강한 양의 상관관계를 보이며, 이는 두 변수 모두 공기 중 화학 물질의 농도 변화를 나타내기 때문이다.
- 미세먼지 농도(PM1.0, PM2.5)와 공기압은 약한 상관관계를 보여주며, 이는 화재 발생 시 공기 중의 입자 수가 변화할 수 있음을 나타낸다.

1) 독립변수와 종속변수 설정

- 독립변수(입력 변수): 온도, 습도, TVOC, eCO2, 센서가 감지한 수소와 에탄올 농도, 공기압, 미세먼지 농도 등 다양한 환경 변수들이 포함됩니다. 이 변수들은 화재 발생 여부에 영향을 줄 수 있는 다양한 요인들을 나타낸다.

- 종속변수(목표 변수): 화재 경보(Fire Alarm). 이 변수는 화재가 발생했는지 여부를 이진(0과 1)으로 나타내며, 예측의 목표가 된다.

4) 러신머닝 선정 이유

- 로지스틱 회귀(Logistic Regression): 이 모델은 이진 분류 문제에 매우 효과적이다. 예측 결과를 해석하기 쉬워, 데이터의 기본적인 패턴을 이해하는 데 유용하다.
- 가우시안 나이브 베이즈(Gaussian Naive Bayes): 연속형 데이터가 정규 분포를 따른다고 가정하여 빠르고 효율적인 예측이 가능하다. 데이터가 비교적 단순하고 변수 간 독립성이 있을 때 유리하다.
- 베르누이 나이브 베이즈(Bernoulli Naive Bayes): 이진 데이터를 처리하는 데 적합하며, 간단한 이진 분류 문제에서 효과적이다.
- 서포트 벡터 머신(Support Vector Machine): 비선형 데이터와 복잡한 분류 문제를 다루는 데 강점이 있으며, 데이터의 경계선을 명확히 하는 데 효과적이다.
- 랜덤 포레스트(Random Forest): 여러 결정 트리를 결합하여 예측 성능을 향상시키는 모델로, 다양한 환경에서의 화재 예측에 적합하다.
- K 최근접 이웃(K-Nearest Neighbors): 데이터 포인트의 근접성을 활용하여 예측하는 간단한 모델로, 데이터의 분포에 대한 사전 가정 없이 사용할 수 있다.
- 익스트림 그라디언트 부스팅(Extreme Gradient Boosting): 여러 개의 간단한 모델들을 결합하여 하나의 강력한 예측 모델을 만드는 알고리즘이다. 이 알고리즘은 여러 모델을 순차적으로 학습시켜 예측 성능을 점진적으로 개선하며, 높은 정확도를 제공한다.
- 인공신경망(Neural Network): 복잡한 데이터 패턴을 학습할 수 있는 모델로, 다양한 환경에서의 화재 예측에 매우 효과적이다.

5) 사용한 성능지표

- 정확도(Accuracy): 전체 데이터 중에서 모델이 올바르게 예측한 비율이다. 전체적인 예측 성능을 평가하는 기본 지표로 사용된다.

- 정밀도(Precision): 모델이 양성(화재 발생)이라고 예측한 것 중에서 실제로 양성인 비율이다. 잘못된 양성 예측을 줄이는 데 중점을 둔다.
- 재현율(Recall): 실제 양성 중에서 모델이 양성으로 올바르게 예측한 비율이다. 실제 양성을 놓치지 않는 데 중점을 둔다.
- F1 점수(F1-Score): 정밀도와 재현율의 조화 평균으로, 두 지표 간의 균형을 평가한다. 두 지표가 모두 중요한 경우 유용하다.

5. 개발 결과

1) 머신러닝 모델의 성능 결과에 대한 해석

- 다양한 AI 모델의 비교 및 평가
 - 로지스틱 회귀, 가우시안 나이브 베이즈, 베르누이 나이브 베이즈, 서포트 벡터 머신, 랜덤 포레스트, K 최근접 이웃, 엑스트림 그라디언트 부스팅 등 여러 분석 모델을 사용하여 데이터를 분석하였다.
- 효과적인 화재 예측 가능성 확인
 - 본 프로젝트에서 사용한 모델들은 대부분 높은 정확도와 정밀도를 보였으며 이는 IoT 장치로부터 수집된 데이터를 바탕으로 화재 발생 가능성을 효과적으로 예측할 수 있음을 의미한다.
- 실제 적용을 위한 잠재력
 - 개발된 AI 모델들은 실시간 연기 감지 및 화재 예측 시스템에 적용할 수 있는 가능성을 보여주었다. 이를 통해 화재 발생 시 신속한 경고를 제공하여 인명과 재산 피해를 최소화 할 수 있을 것으로 기대된다.

6. 결론

1) 기대 효과

- IoT(사물인터넷) 장치로부터 수집된 데이터를 활용하여 AI 기반 연기 감지 모델을 개발하고, 이를 통해 보다 스마트하고 빠르게 화재를 예측할 수 있는 시스템을 구축할 수 있다.
- 화재를 조기에 감지하고 신속하게 대응함으로써 인명과 재산 피해를 최소화하고 공공 안전을 증진한다.

2) 머신러닝 모델의 한계

- 데이터 다양성의 부족
 - 현재 모델은 특정 환경과 조건에서 수집된 데이터로 훈련되었다. 다양한 환경에서의 데이터가 필요한 상태이다.
- 모델 복잡성 및 실행시간
 - 일부 고성능 모델, 특히 인공신경망은 많은 계산 자원을 필요로 하며, 실시간 예측에 속도개선이 필요하다.
- 센서 데이터의 신뢰성
 - IoT 장치에서 수집된 센서 데이터의 정확도가 모델 성능에 직접적인 영향을 미친다. 센서의 정확도와 유지 보수가 중요하다.
- 환경변수의 변화
 - 온도, 습도 등 환경 변수는 시간과 조건에 따라 변동성이 크다. 따라서 이를 효과적으로 반영하기 위한 데이터 업데이트가 필요하다.