

# Data-Driven Smoke Detection

이름: 임경찬

학번: 1918077

Github:

## 1. Data-Driven Smoke Detection 모델 개발 관련 요약

데이터를 중심으로 분석하여 정확한 화재감지률을 95% 이상으로 하여, 수많은 화재감지기의 오작동률을 줄이는 것이 주 목적이다. 왜냐하면 수많은 기업, 또는 아파트 지하주차장에서 화재감지기의 오작동으로 인해 운영 및 생산에 차질이 생긴다는 이유로 임의로 화재감지기의 전원을 꺼두는 만행을 하고 있기에, 이를 예방하기 위해서 이다.

## 2. 개발 목적

### a. Data-Driven Smoke Detection 모델 활용 대상:

1. 아파트 지하주차장 화재감지기
2. 제조업 및 건설업 등 다양한 분야에서의 화재감지기
3. 전기자동차 내부에서의 화재감지기 ( 화재 감지 시 전기차 전원 정지 및 긴급상황 알림 )

### b. 개발의 의의:

#### 1. 정확한 위험 분석:

다양한 환경 데이터(예: 온도, 습도, 공기 질 데이터 등)를 기반으로 화재 발생 가능성을 분석하여 오탐(false positive)과 미탐(false negative)을 줄일 수 있다.

#### 2. 지속 가능한 발전:

데이터 기반의 화재 감지 모델은 다양한 환경 변화에 맞춰 학습이 가능하여, 향후 지속적으로 성능을 개선할 수 있는 확장 가능성을 가진다.

### c. 데이터의 독립 변수와 종속 변수 :

#### 1. 독립변수

- Temperature (온도)
- Humidity (습도)
- TVOC (Total Volatile Organic Compounds, 총휘발성유기화합물)
- eCO2 (Equivalent Carbon Dioxide, 등가 이산화탄소): 공기 중 이산화탄소의 농도
- PM1.0, PM2.5, PM10: 공기 중 미세먼지 입자의 농도
- NC0.5, NC1.0, NC2.5: 특정 크기의 입자 농도
- H<sub>2</sub> (수소 농도) 및 CH<sub>4</sub> (메탄 농도)
- Pressure (압력) : 대기압

#### 2. 종속변수

- Fire Alarm

화재가 발생하지 않은 경우: 0

화재가 발생한 경우: 1

## 3. 배경지식

### a. 데이터 관련 사회 문제

지난 2024년 8월 1일 인천 청라 지하주차장에서 전기차 화재로 인하여 약 959대 차량 및 시설물 피해가 발생하여 피해 규모액만 약 38억에 달했다.

이렇게까지 피해규모가 커진 이유 중 하나는 스프링클러가 작동하지 않았기 때문이다. 화재 감지 센서는 정상적으로 작동하였지만 스프링클러 밸브가 잠겨있어 물이 전달되지 못했다.

경찰 조사 결과, 스프링클러 밸브를 잠근 이유는 수시로 화재감지가 울려 스프링클러가 작동되면 수많은 민원이 발생하기에 이를 방지하고자 잠근 것으로 확인되었다.<sup>1</sup>

결과적으로 이번 사건에서는 화재 감지가 정상적으로 작동하였지만 과거에 정확도가 낮아서 오작동이 빈번하게 발생하여 스프링클러 밸브를 잠근 것이기에, 화재감지기의 정확도가 높은 모델의 개발이 시급하다.

---

<sup>1</sup> 인천투데이, <https://www.incheontoday.com/>, <청라 벤츠 전기차 화재 참사...스프링클러 평소에도 오작동 경고>, 김현철 기자, 2024-08-12.

## 4. 개발 내용

### a. 데이터에 대한 구체적 설명

- 데이터 개수

행의 개수 : 62,253개

열의 개수 : 15개 (fire alarm 포함)

- 데이터 속성

Temperature 주변 환경의 온도 섭씨 [°C]

Humidity 주변 환경의 상대 습도 [ % ]

TVOC 총휘발성유기화합물 농도 [ppb (parts per billion) ]

eCO<sub>2</sub> 등가 이산화탄소 농도 [ppm (parts per million)]

H<sub>2</sub> 수소 농도 [ ppm ]

CH<sub>4</sub> 메탄 농도 [ ppm ]

PM1.0 / PM2.5 / PM10 공기 중의 미세먼지 입자 농도 [ µg/m<sup>3</sup> ]

NC0.5 / NC1.0 / NC2.5 특정 크기의 입자 농도 [ 개/리터 ]

Pressure 대기압 [hPa]

Fire Alarm 화재 여부를 나타내는 이진 변수 (0: 화재 아님, 1: 화재)

### b. Data-Driven Smoke Detection 모델의 목적

2-c에서 설정한 독립변수와 종속변수와의 관계를 데이터를 통해서 분석하고 어떤 요소가 화재감지의 오작동을 유발하는지, 그리고 어떤 변수가 화재감지의 정확도를 높이는지 알아보고 오작동률이 낮고 정확도가 높은 화재감지기 모델을 만드는 것이 최종 목적이다.

### c. 사용한 성능 지표

1. 정확도(Accuracy) : 전체 예측에서 올바르게 예측한 데이터 비율이다.
2. F1 Score : 정밀도와 재현율의 조화 평균으로, 특히 클래스 불균형이 있는 경우 유용한 지표이다.
3. AUC(ROC AUC) : ROC 곡선 아래의 면적을 기반으로 모델이 양성과 음성을 얼마나 잘 구별하는지 평가하기 위한 지표이다.

이 세 가지 지표는 모델이 얼마나 잘 분류하고 예측하는지 전반적으로 평가할 수 있는 지표들이라고 판단하였고, 머신러닝 모델의 성능을 종합적으로 분석하는 데 적합하다고 생각하여 이 지표들을 선정했다.

## 5. 개발 결과

### a. 성능 지표에 따른 머신러닝 모델 성능 평가

본 프로젝트에서는 Accuracy, F1 Score, ROC AUC와 같은 주요 성능 지표를 사용하여 모델의 성능을 평가하였다. 다음은 최종 모델의 평가 결과이다.

- 1) Accuracy : 1.0000
- 2) F1 Score : 1.0000
- 3) ROC AUC : 1.0000

모델이 모든 예측을 정확히 수행하였음을 나타내며, 분류 성능이 매우 우수함을 확인할 수 있다.

### b. 머신러닝 모델의 성능 결과에 대한 해석

#### 1. Classification Report

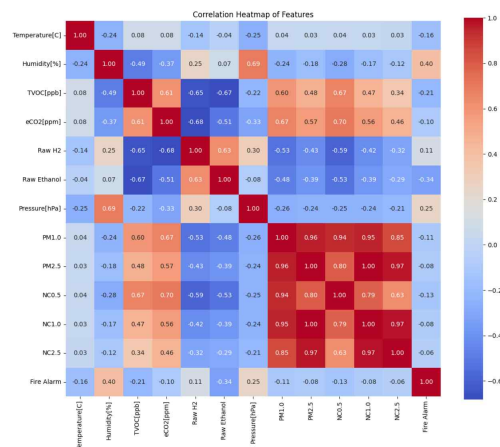
Class	Precision	Recall	F1-Score	Support
0 (No Fire)	1.00	1.00	1.00	17,873
1 (Fire)	1.00	1.00	1.00	44,757
Accuracy			1.00	62,630
Macro Avg	1.00	1.00	1.00	62,630
Weighted Avg	1.00	1.00	1.00	62,630

모델은 화재 탐지 (No Fire or Fire)에 대해 모든 데이터를 완벽하게 예측하였다.

그리고 Precision, Recall, F1-Score 모두 1.00으로 모델 성능이 이상적으로 동작하고 있는 것을 알 수 있다.

하지만, 현실 데이터에 적용할 경우 과적합가능성이 있으므로 추가 검증이 필요해 보인다.

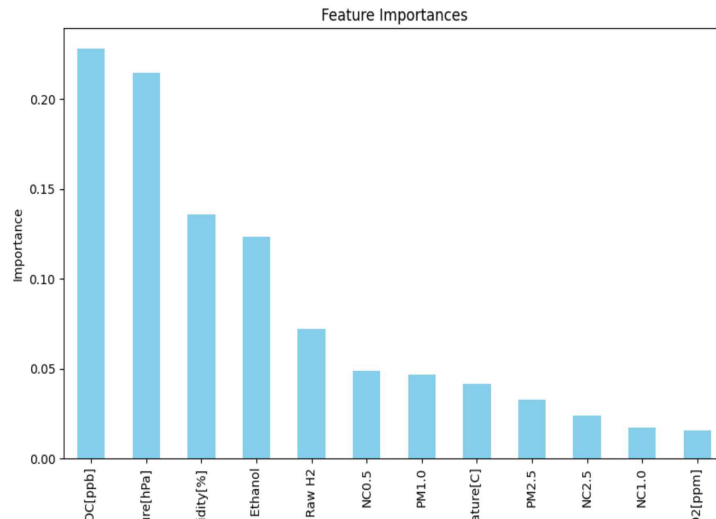
#### 2. 상관 관계 히트맵



히트맵에서 발견되는 부분은 Fire Alarm 과 습도의 관계가 0.40으로 가장 높은 양의 값이 나왔다.

따라서 습도가 높을수록 화재 알림이 울릴 확률이 높다는 것을 알았고, 이를 다시 해석하면 화재 알림의 오작동률의 주요 변수는 습도라는 것을 알 수 있다.

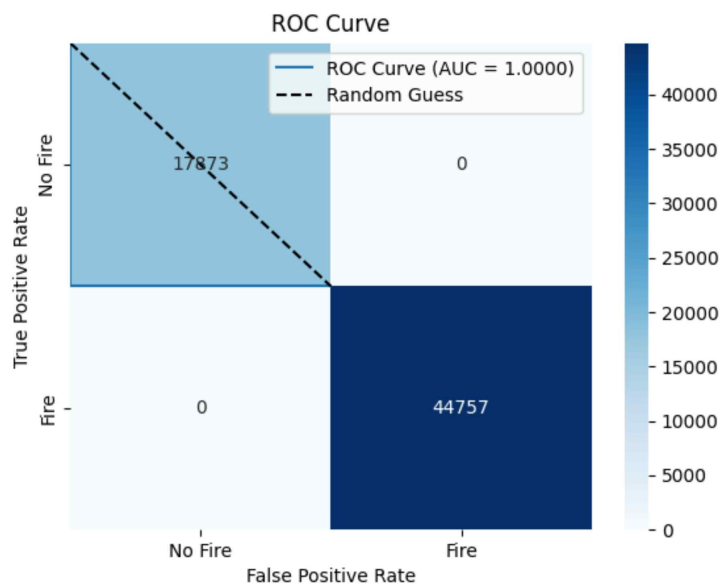
### 3. 중요도



TVOC[ppb]와 Pressure[hPa]의 변수가 가장 다른 변수와 다르게 가장 중요한 변수임을 알 수 있다.

따라서 이 두 변수를 중심으로 센서 개선, 데이터 수집 최적화, 경보 시스템 강화를 통해 더욱 높은 신뢰도를 가진 화재 감지 시스템을 구축할 수 있다는 것을 알 수 있다.

### 4. ROC Curve



True Negative (No Fire, 예측 0 → 실제 0): 17,873건

True Positive (Fire, 예측 1 → 실제 1): 44,757건

False Positive (No Fire, 예측 1 → 실제 0): 0건

False Negative (Fire, 예측 0 → 실제 1): 0건

이 모델은 모든 데이터를 완벽하게 분류했으며 오류가 전혀 발생하지 않았음을 알 수 있다. 따라서 분석 결과 모델은 완벽한 성능을 달성했지만 과적합 여부를 고려하여 추가적인 검증이 필요해보인다.

## 6. 결론

### a. 머신러닝 모델 개발에 관한 간략한 요약 및 결과 설명

본 프로젝트에서는 화재 탐지 모델을 개발하였으며, 최종적으로 Accuracy, F1 Score, ROC AUC 모두 1.000의 성능을 기록하였다. 그리고 상관관계 분석에서 습도가 화재 알림에 중요한 변수임을 알았고, 중요도 분석을 통해 TVOC와 대기압이 주요한 영향 변수임을 확인할 수 있다.

### b. 개발 의의

2-b에서 언급했듯이 개발의의는 정확한 위험 분석과 지속 가능한 발전이다.

따라서 이번 프로젝트를 통해서 화재 알림을 개발할 때에는 오작동을 유발하는 습도를 조절할 수 있는 장치가 필요하며, TVOC(총유기화합물)과 대기압을 중심으로 센서 개선, 데이터 수집 최적화, 경보 시스템 강화를 통해 더욱 높은 신뢰도를 가진 화재 감지 시스템을 만들어야 한다는 것을 알 수 있다.

### c. 머신러닝 모델의 한계

이 모델은 모든 데이터를 완벽하게 분류했으며 오류가 전혀 발생하지 않았음을 알 수 있었는데 오히려 오류가 없었다는 점에서 한계점이 있다고 생각한다.

과적합 문제가 있다고 판단되며 이는 추가로 테스트가 필요하다고 생각된다. 그리고 다양한 환경과 조건에서의 데이터가 부족하다고 판단되며 그렇기에 더 다양한 데이터가 필요하다고 생각한다.

### d. 느낀점

솔직히 이번 프로젝트를 맡았을 때 정말 어떻게 해야 할지 막막했으며, 프로젝트를 완성할 자신이 전혀 없었다. 하지만 데이터 분석과 머신러닝 모델 개발 과정을 차근차근 배우고 직접 실행하면서 점차 자신감을 얻을 수 있었다.

처음에는 복잡한 데이터셋과 성능 평가 지표에 대한 이해 부족해 진행 자체가 어려웠다. 하지만 문제를 단계별로 나누고 한 가지씩 해결해 나가면서 데이터 전처리, 시각화, 모델

학습 및 평가까지 마무리할 수 있었다.

특히 프로젝트 결과에서 모델의 성능이 완벽에 가까운 수치를 기록한 순간 그 동안의 고민과 노력에 대한 보상을 받는 것 같아 정말 기뻐다. 물론 이러한 수치가 비정상적이기에 또 무언가 잘못했겠다는 생각도 들었지만 이번 프로젝트를 통해 데이터 분석 능력뿐만 아니라 문제 해결력과 끈기를 키울 수 있었다고 생각했고 이러한 경험은 앞으로의 도전에 큰 도움이 될 것이라 확신한다.