

LPG 가스 누출 탐지 시스템 - LGG

발표자 : 김민석

팀원 : 김민석, 이승연, 임하영, 장영수

Github : -

1. LPG 가스 누출 탐지 프로그램 개요

- a. LPG(액화석유가스)는 가정과 산업에서 광범위하게 사용되지만, 가스 누출 시 심각한 화재와 폭발 사고를 초래할 수 있다. 최근 LPG 가스 누출로 인한 사고가 빈번하게 발생하면서, 이를 예방하고 안전성을 높이기 위한 시스템 개발의 필요성이 커지고 있다.

2. 개발 목적

- a. 개발의 의의 : 다양한 화학물질을 다루는 산업현장에서 사용할 수 있도록 Python 기반의 시스템을 개발하였으며, 데이터를 분석하여 LPG 가스 누출 여부를 이진법으로 판별해 사전에 화재를 예방한다.

3. 배경지식

- a. LPG 가스 : 연료와 화학 연료로 널리 사용되며, 산업현장, 상업시설, 가정 등 다양한 분야에서 사용한다. 무색,무취로 일반적으로 감지하기 어렵기 때문에 누출이 발생할 경우 조기에 발견하지 못하면 대규모 폭발 및 화재 야기한다.
- b. 머신 러닝(Machine Learning) : 종속 변수와 하나 이상의 독립 변수 간의 관계를 설명하고, 예측하는데 사용한다.
- c. 회귀분석(Regression Analysis) : 데이터 간의 관계를 모델링하고 학습함으로써 미래의 값을 예측하고, 특정 변수가 다른 변수에 미치는 영향을 파악하는 통계적 기법이다.
- d. 선형 회귀(Linear Regression) : 독립 변수와 종속 변수 간의 관계를 선형으로 모델링한다.

4. 개발 내용

a. 데이터에 대한 설명 및 시각화 정보

- i. 데이터의 독립변수(속성) : 8 가지의 각종 가스 및 요인 (Alcohol, CH₄, CO, H₂, LPG, Propane, Smoke, Temp)
- ii. 데이터의 종속변수 : LPG 가스 누출 여부(LPG-Leakage)
- iii. 총 데이터 개수 1000 개
- iv. 총 8 가지의 가스와 온도 데이터를 활용
 1. Alcohol : 높은 농도로 존재하면, 다른 가스들과 혼합될 때 그 상호작용이 LPG 측정에 영향을 줄 수 있다.
 2. CH₄ : 메탄은 LPG 는 모두 탄화수소 계열로, 메탄 농도가 높으면 LPG 및 가스 혼합물의 폭발에 영향을 줄 수 있다.
 3. CO : 일산화탄소의 농도가 높다면 LPG 및 다른 가스의 연소를 야기한다.
 4. H₂ : 수소와 LPG 가 동시에 존재하는 경우, 두 가스 모두 폭발 위험을 야기한다.
 5. LPG : 프로판과 부탄의 혼합물로 상온에서 액체 상태로 저장되지만, 누출 시 기체 상태로 변해 공기 중으로 퍼진다.
 6. Propane : 프로판은 LPG 의 주요 성분으로 두 값이 비례관계를 가져 프로판의 농도가 높으면 LPG 의 농도도 높을 가능성이 높다.
 7. Smoke : 불완전한 연소를 나타내며, 연기가 감지되면 LPG 또는 다른 연료가 제대로 연소되지 않고 있음을 나타냄. 즉, 연기의 존재는 타 가스의 연소를 야기한다.
 8. Temp : 온도가 높으면 LPG 가 기화하여 공기 중에 퍼질 가능성이 증가한다.
- v. 데이터를 수집 및 전처리해 LPG 가스의 누출 여부를 확인
 1. 데이터 수집 : 데이터는 csv 파일로 되어있고, 8 개의 독립변수에 관한 데이터 1000 개를 바탕으로 가스 및 온도 측정값으로 LPG 가스의 누출 여부 확인한다.
 2. 데이터 전처리 : 데이터 분석과 모델 학습의 품질을 높이기 위해 데이터를 정리하고 변환하는 과정으로 다음과 같은 순서로 실행했다.

3. 데이터 준비 - array 를 사용해 8 개의 독립변수를 가진 데이터를 배열(list)형태로 표현한다.
4. 데이터 정규화 – MinMaxScaler 를 사용해 1000 개의 데이터를 0 과 1 사이의 값으로 조정해 모든 데이터를 일정한 범위로 변환한다.

b. 머신 러닝 모델 선정 이유

	A	B	C	D	E	F	G	H	I	J
1	Alcohol	CH4	CO	H2	LPG	Propane	Smoke	Temp	LPG_Leakage	
2	0.002335	3.047929	3.151109	18.36901	0.020111	0.007674	0.021269	31	1	
3	0.003798	2.084298	2.771445	26.24979	0.011426	0.008733	0.023268	37	1	
4	0.004403	4.325731	4.675765	23.5681	0.060044	0.008404	0.017806	36	1	
5	0.004409	2.384824	3.655571	19.9225	0.088107	0.008073	0.027534	26	0	
6	0.002766	5.026507	3.751914	24.5239	0.074812	0.009797	0.027073	35	1	
7	0.004292	5.570761	4.522306	18.8231	0.045346	0.00887	0.020419	27	0	
8	0.003814	2.992165	0.47974	11.80377	0.039146	0.006067	0.020664	35	0	
9	0.003168	4.803138	5.713318	22.17844	0.066526	0.006757	0.018238	30	0	
10	0.004191	1.684397	3.930553	9.606966	0.063133	0.00814	0.022734	32	1	

i. 해당 데이터를 판단하기 위한 머신 러닝 모델 선정

1. 로지스틱 회귀(Logistic Regression) 모델 선정 :

해당 데이터의 종속 변수(0 or 1)와 같은 이진분류 문제를 해결하는 데 사용하는 통계적 모델로 LPG 가스의 누출 또는 정상을 판단하기에 적합하다고 생각해 사용하였다.

ii. 성능 비교를 위한 머신 러닝 모델 선정

1. 의사결정 나무(DecisionTreeClassifier) 모델 선정 :

해당 데이터의 독립변수 값처럼 비 선형적이고 복잡한 데이터 패턴을 효과적으로 모델링할 수 있다. 다양한 특성 간의 비선형 관계를 포착해 데이터의 복잡성을 반영하는 것이 단순 데이터를 판단하는데 유리한 로지스틱 회귀 모델과 비교를 위해 사용하였다.

c. 머신 러닝 성능 지표

i. 사용한 성능 지표 종류

1. 정확도(Accuracy)

- a. 전체 데이터 중에서 올바르게 분류된 샘플의 비교를 나타내는 성능 지표로 모델의 전반적인 성능을 간단명료하게 이해할 수 있어서 사용하였다.

2. ROC-Curve

- a. 원래는 신호 감지 이론에서 사용되던 개념이지만 현재는 머신 러닝의 성능을 평가하는데 사용된다. 다양한 임계 값에 대한 성능을 시각화 하여 곡선이 왼쪽 상단에 가까울수록 모델의 성능이 좋다는 것을 한눈에 판단 가능해서 사용하였다.

- ii. 성능 지표 선정 이유

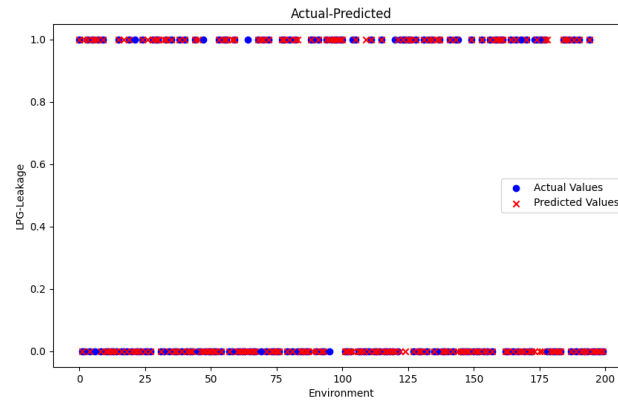
- 1. 머신 러닝 모델이 LPG 가스 누출 상황을 정확하게 판단하는 것이 가장 중요하다. 만약 모델이 가스 누출이 발생했음에도 불구하고 이를 감지하지 못하거나 ‘없다’고 판단하면, 심각한 안전 위험을 초래할 수 있다.
- 2. 모델의 정확도가 높아야만 이러한 위험을 최소화하고 사고를 사전에 예방할 수 있다고 생각하여 ‘Accuracy’, ‘ROC-Curve’ 성능 지표를 선정해 모델이 얼마나 신뢰성 있게 가스 누출을 감지하는지 평가하고자 하였다.

5. 개발 결과

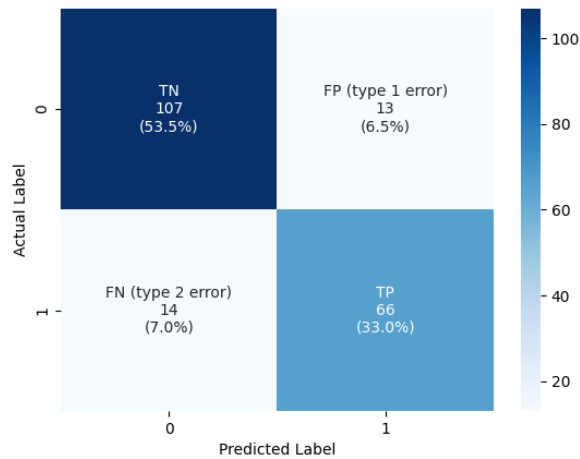
a. LPG 가스 누출 탐지 모델 머신 러닝 결과

i. 로지스틱 회귀 모델 적용 시 결과

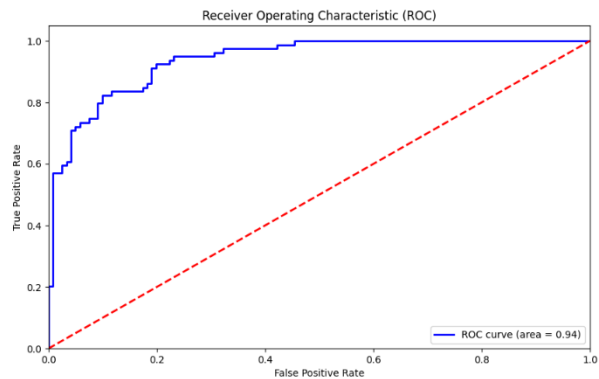
1. 머신 러닝 결과



2. 머신 러닝 성능지표 Accuracy 결과값

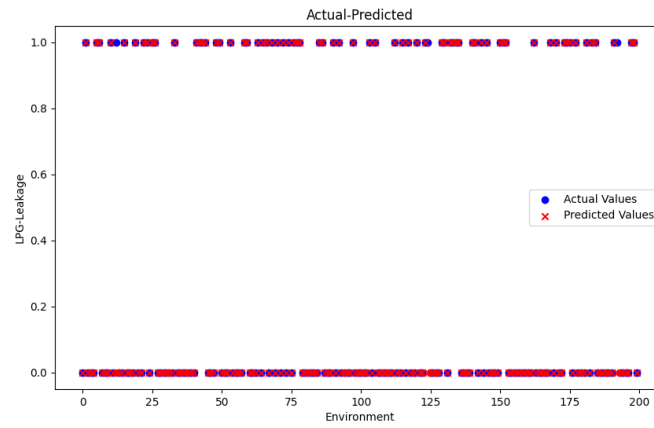


3. 머신 러닝 성능지표 ROC-Curve 결과값

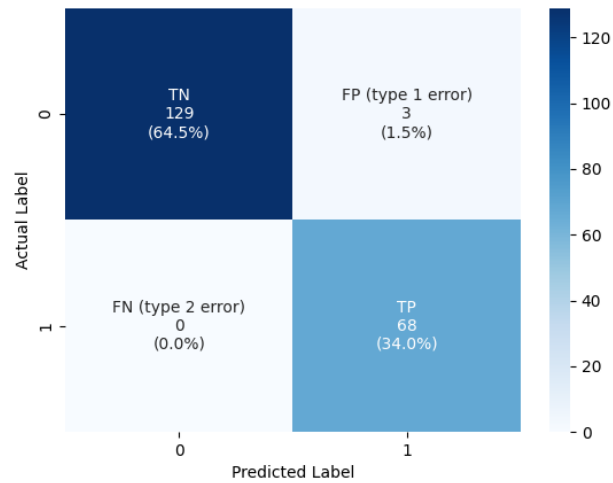


ii. 의사결정 나무 모델 적용 시 결과

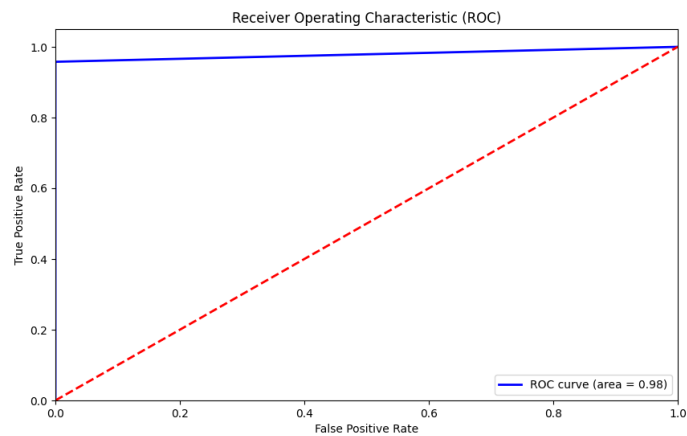
1. 머신 러닝 결과



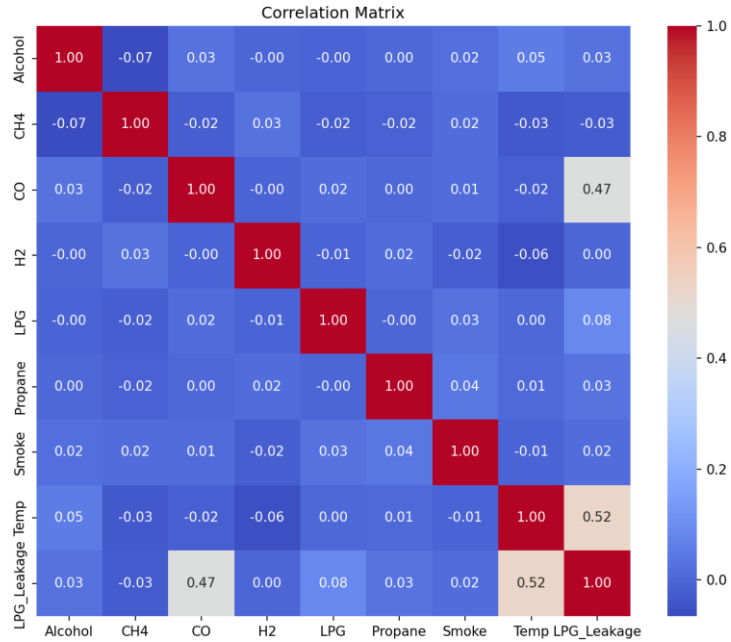
2. 머신 러닝 성능지표 Accuracy 결과값



3. 머신 러닝 성능지표 ROC-Curve 결과값



iii. HEATMAP 을 통한 독립변수와 종속변수의 상관관계 분석



iv. 머신 러닝 모델 간 성능 비교

- 로지스틱 회귀 모델, 의사결정 나무 모델을 사용했을 시 각각의 머신 러닝 결과를 확인하였다. 그 후 성능 지표를 활용해 교차 검증을 하여 결과 정확도를 비교해보았다.
- 로지스틱 회귀 모델의 정확도는 **0.8480000000000001**
- 의사결정 나무 모델의 정확도는 **0.9970000000000001**

b. 머신 러닝 모델의 성능 결과에 대한 해석

- 두 모델의 머신 러닝 결과만을 봤을 때는 차이를 확인하기 어려웠으나 성능 지표를 교차 검증으로 두 머신 러닝 간 정확도 차이를 확인할 수 있었고, 의사결정 나무 모델이 로지스틱 회귀 모델에 비해 월등한 정확도를 가지고 있다는 결과를 도출하였다.
- HEATMAP 의 시각화 결과 온도(Temp)와 일산화탄소(CO)가 LPG 가스 누출에 가장 큰 영향을 끼치는 것을 알 수 있었다.

6. 결론

a. 결론

- i. LPG 가스 누출 탐지를 위한 의사결정 나무 머신 러닝 모델과 그 결과에 대한 성능 지표 분석은 100%에 가까운 정확도를 도출해냈다. 현재는 1000 개의 가상 데이터를 분석하였으나 실제 현장에서는 유해가스탐지기 등을 이용해 실시간으로 정보를 입력 받고 그 정보를 바탕으로 LPG 가스 누출을 탐지할 수 있는 프로그램으로 응용 가능할 것이다.
- ii. LPG 가스 누출과 큰 상관관계를 가진 온도(Temp)와 일산화탄소(CO) 만으로도 LPG 가스 누출 여부 확인에 대해 알 수 있을 것이다.