

LGG(LPG Gas Guard)

Early-LPG-Gas-leakage-prediction-system

6조 - 김민석, 이승연, 임하영 장영수

Github - minseok3/team6_pjt- (github.com)



개요 및 개발 목적

01 개요

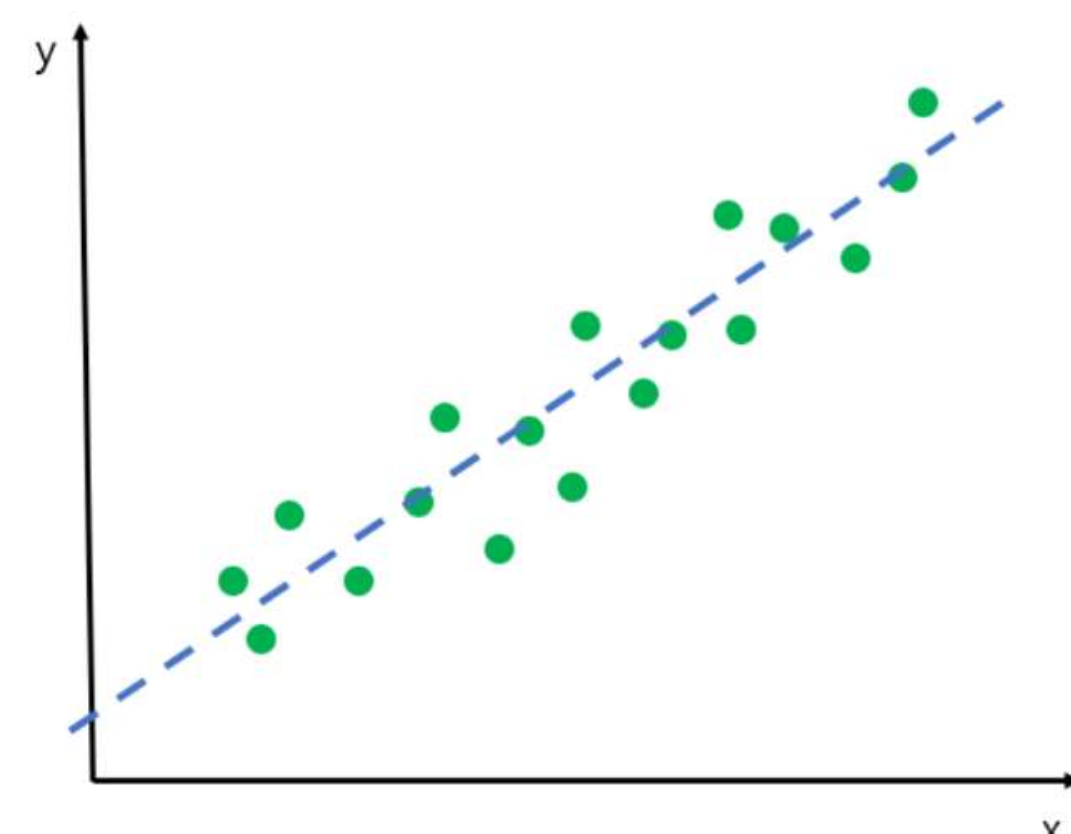
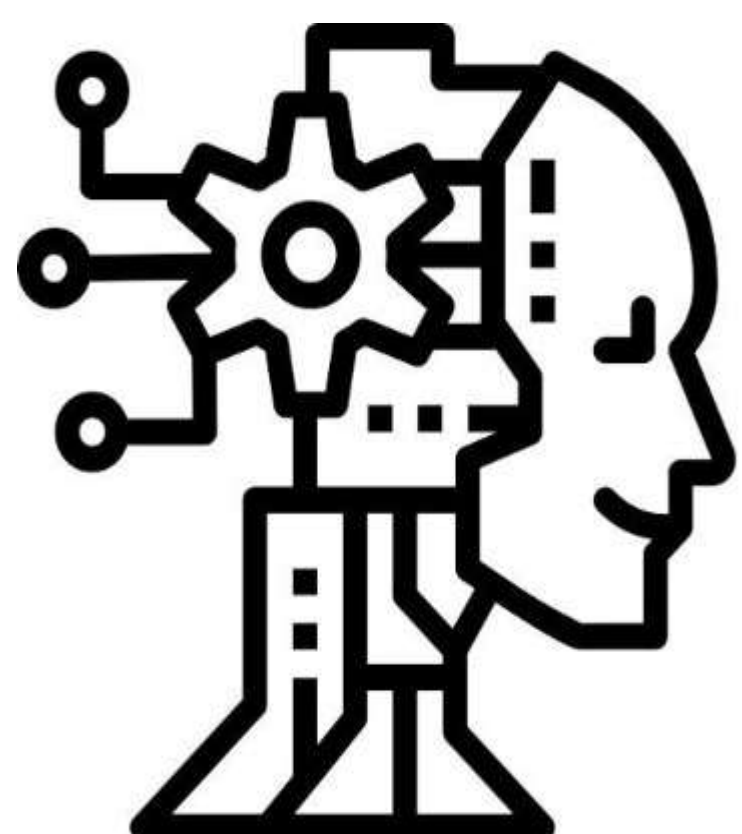
LPG(액화석유가스)는 가정과 산업에서 광범위하게 사용되지만, 가스 누출 시 심각한 화재와 폭발 사고를 초래. 최근 LPG 가스 누출로 인한 사고가 빈번하게 발생하면서, 이를 예방하고 안전성을 높이기 위한 시스템 개발의 필요성이 커짐에 따라 우리는 로지스틱 회귀 모델과 의사결정 나무 모델을 활용해 LPG가스 누출 여부를 정확하게 판단하는 프로그램을 개발하여

가스 누출의 조기 탐지를 신속한 대응을 통해 안전성 확보.

02 개발 목적

본 프로그램의 주요 목적은 LPG가스 누출을 신속하고 정확하게 탐지하여 안전사고를 예방하는 것. LPG가스는 무색, 무취인 특성으로 인해 누출 시 감지가 어렵고, 이로 인한 심각한 화재와 폭발 사고 발생 야기.

로지스틱 회귀 모델과 의사결정 나무 모델을 활용하여 가스 누출의 가능성을 신속하게 분석하고 조기에 탐지. 다양한 데이터 입력 값을 바탕으로 LPG가스 누출 여부를 정확하게 판단하여 오탐지와 탐지 누락을 최소화. 가정과 산업 현장에서의 안전성을 높이고, 가스 누출로 인한 화재 및 폭발사고를 예방하여 인명과 재산을 보호. 실시간으로 데이터를 분석하고 결과를 제공하여, 신속한 대응과 문제 해결을 지원.



배경지식 및 사고 사례

01 배경지식

- LPG 가스는 산업, 상업, 가정 등에서 사용되는 무색, 무취의 연료로 누출 시 대규모 폭발이나 화재를 일으킬 수 있어 감지 장치가 필요.
- 머신 러닝은 독립 변수와 종속 변수 간의 관계를 학습하여 예측하는 기술.
- 회귀분석은 데이터 간의 관계를 모델링하여 미래 값을 예측하고 변수 간의 영향을 분석하는 통계 기법.
- 선형회귀는 독립 변수와 종속 변수 간의 관계를 선형으로 모델링.

02 사고 사례 - 평창 LPG가스 충전소 폭발 사고

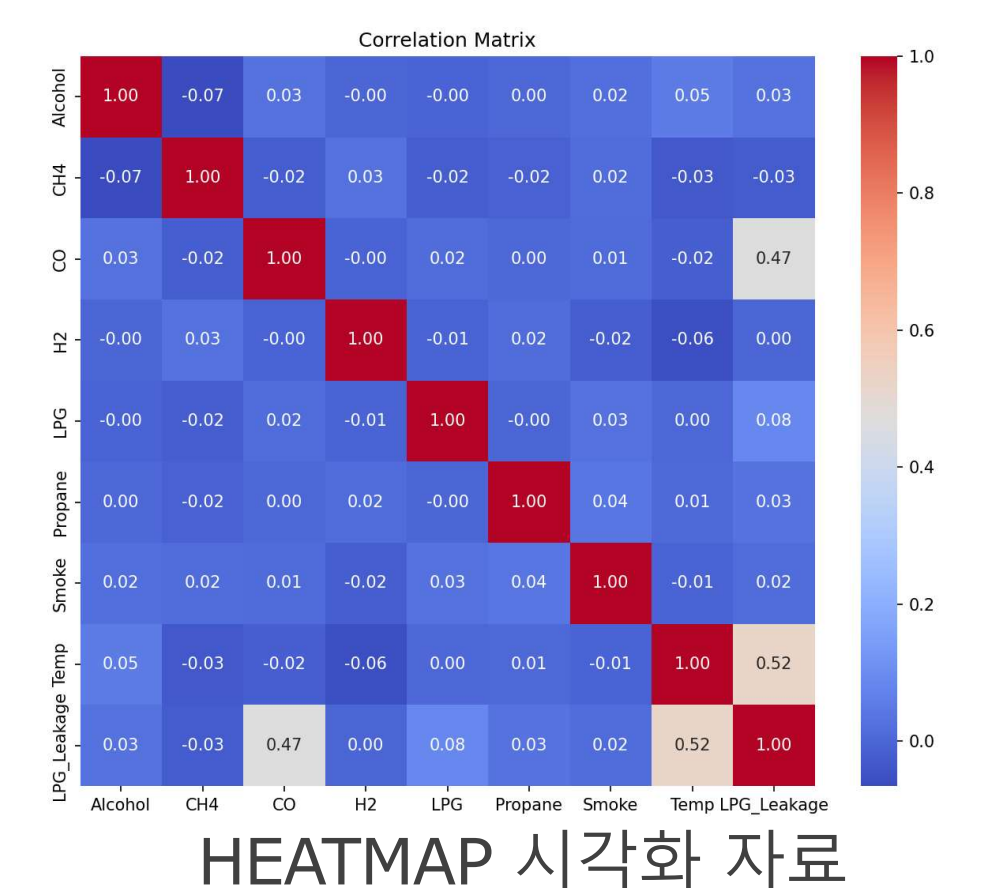
- 2024년 1월 1일 오후 8시 40분 경 강원도 평창군 용평면 장평리에서 LPG가스가 새고 있다는 신고가 접수되었고, 20분 뒤 폭발사고 발생.
- 벨크로리를 통해 가스 이.충전하는 과정에서 LPG가스가 누출되어 주변으로 확산 되었고, 원인미상 점화 원인에 의해 폭발 -> 주민 5명 부상, 1명 사망, 주택 14채 파손, 차량 14대 피해

- LPG가스 취급 시설의 안전 점검 및 유지보수 강화와 가스 누출 감지 시스템의 설치 필요.

개발 내용 - 상관관계 분석

01 HEATMAP을 통한 데이터간 상관관계 분석

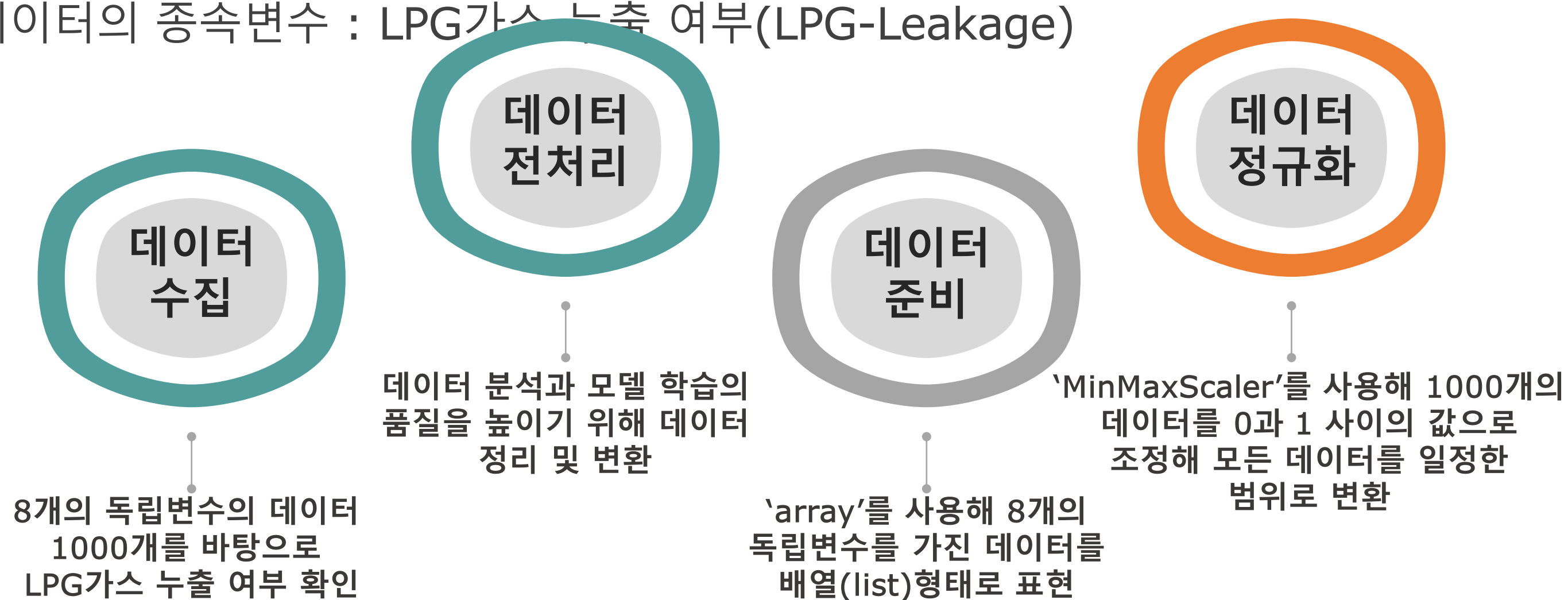
- LPG가스 누출 탐지에 가장 큰 영향을 미치는 변수 값이 Temp, CO라는 것을 파악.
- 유해가스 측정기가 여러 개의 가스 값을 측정하고, 현장 상황에 맞추어 두가지 변수 값만으로는 LPG가스 누출 여부를 확인하기에는 무리가 있다고 판단, 6개 가스 값을 추가하여 8개 가스와 요소를 통한 데이터 분석을 진행.



개발 내용 - 데이터 분석 및 시각화

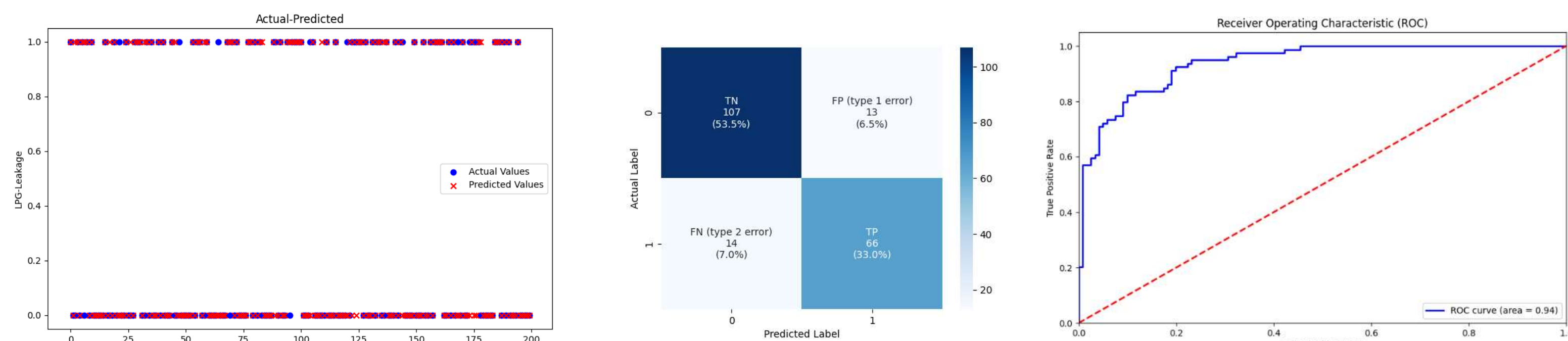
01 데이터에 대한 설명

- 데이터의 독립변수(속성) : 농도가 높다면 LPG 및 다른 가스의 연소를 야기하는 일산화탄소(CO), 기온이 올라가면 LPG가 기화하여 공기 중에 퍼질 가능성이 커지는 온도(Temp) 외 6가지의 추가적인 각종 가스(Alcohol, CH4, H2, LPG, Propane, Smoke)
- 데이터의 종속변수 : LPG가스 누출 여부(LPG-Leakage)



04 로지스틱 회귀 model

- 해당 데이터의 종속 변수(0 or 1)와 같은 이진 분류 문제를 해결하는 데 사용하는 통계적
- 모델로 LPG가스의 '누출' 또는 '누출되지 않음'을 판단하기에 적합하다고 생각해서 사용.



02 사용한 성능 지표 종류

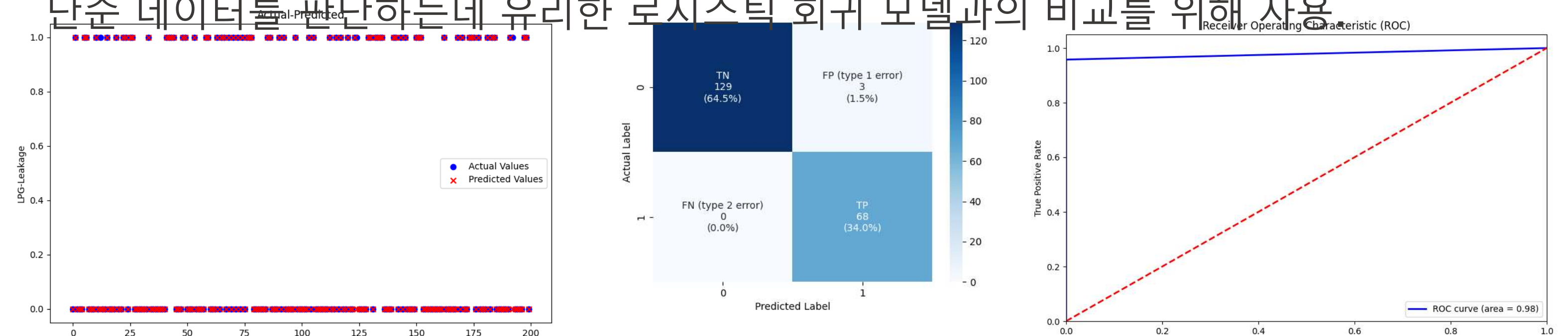
- 정확도(Accuracy)
 - 전체 데이터 중에서 올바르게 분류된 샘플의 비교를 나타내는 성능 지표로 모델의 전반적인
 - 성능을 간단명료하게 이해할 수 있어 사용.
- ROC Curve
 - 원래는 신호 감지 이론에서 사용되던 개념이지만 현재는 머신 러닝의 성능을 평가하는데 사용되고, 다양한 임계 값에 대한 성능을 시각화 하여 곡선이 왼쪽
 - 상단에 위치할수록 모델의 성능이 좋다는 것을 한눈에 판단 가능해서 사용.

03 성능 지표 선정 이유

- 머신 러닝 모델이 LPG가스 누출 상황을 정확하게 판단하는 것이 가장 중요.
- 만약 모델이 가스 누출이 발생했음에도 불구하고 이를 감지하지 못하거나 '없다'라고 판단하면, 심각한 안전 위험을 초래.
- 모델의 정확도가 높아야 이러한 위험을 최소화하고 사고를 사전에 예방할 수 있다고 생각해 두 성능 지표를 선정해 모델의 가스 누출 감지 신뢰성을 평가.

05 의사결정 나무 model

- 해당 데이터의 독립변수 값처럼 비선형 적이고 복잡한 데이터 패턴을 효과적으로 모델링 가능하고, 다양한 특성 간의 비선형 관계를 포착해 데이터의 복잡성을 반영하는 것이
- 단순 데이터를 판단하는데 유리한 로지스틱 회귀 모델과의 비교를 위해 사용.



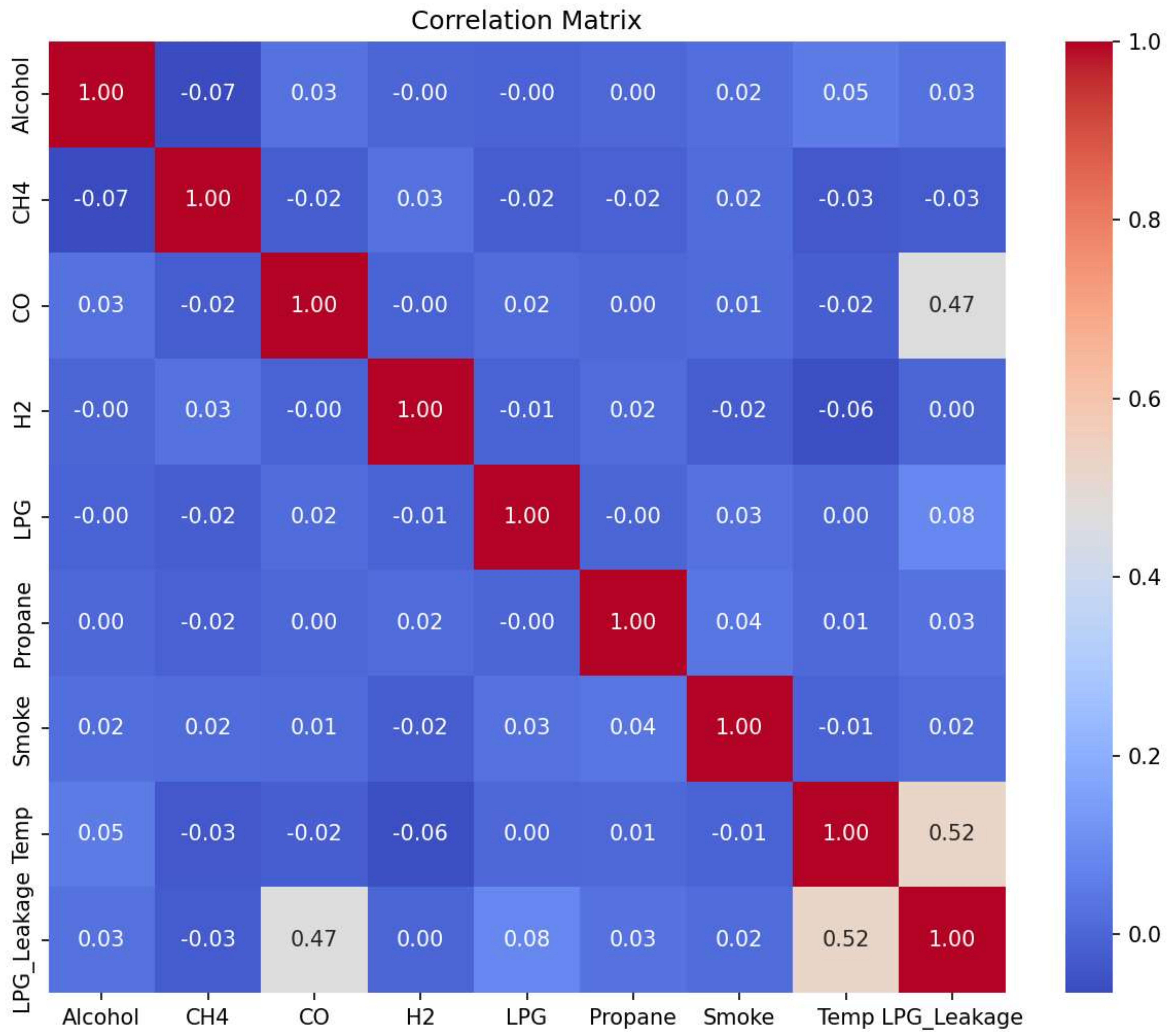
최종 개발결과 및 결론

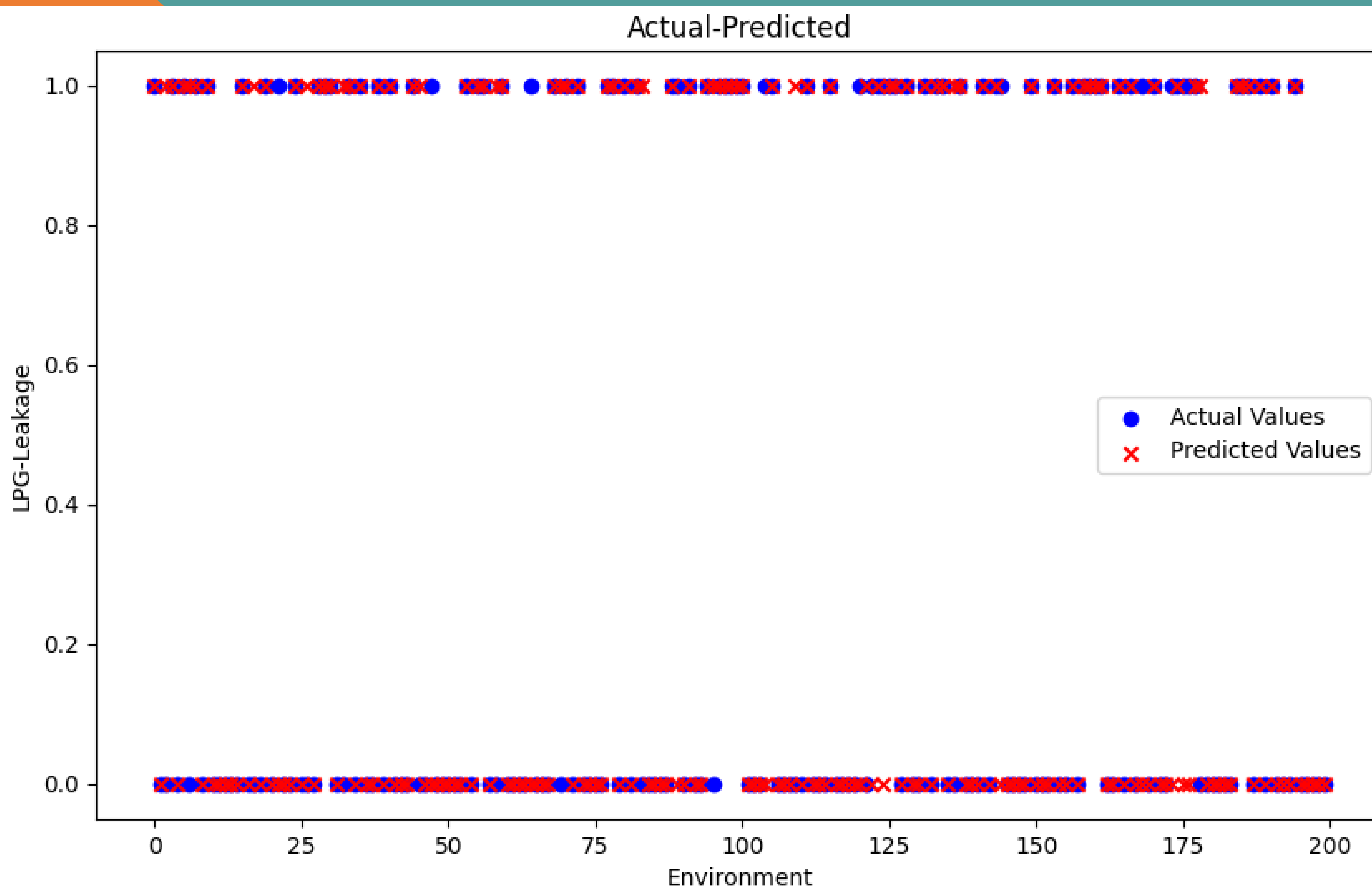
01 최종 개발결과

- 8개의 독립변수 당 800개의 학습 데이터와 200개의 테스트 데이터를 사용하여 두 모델의 머신 러닝 결과만을 봤을 때는 차이를 확인하기 어려웠으나, 성능 지표를 교차 검증해 두 머신 러닝 간 정확도가 15% 차이를 보여 유의미한 결과를 확인. 따라서 의사결정 나무 모델이 로지스틱 회귀 모델에 비해 월등한 정확도를 가지고 있다는 결과를 도출.

02 결론

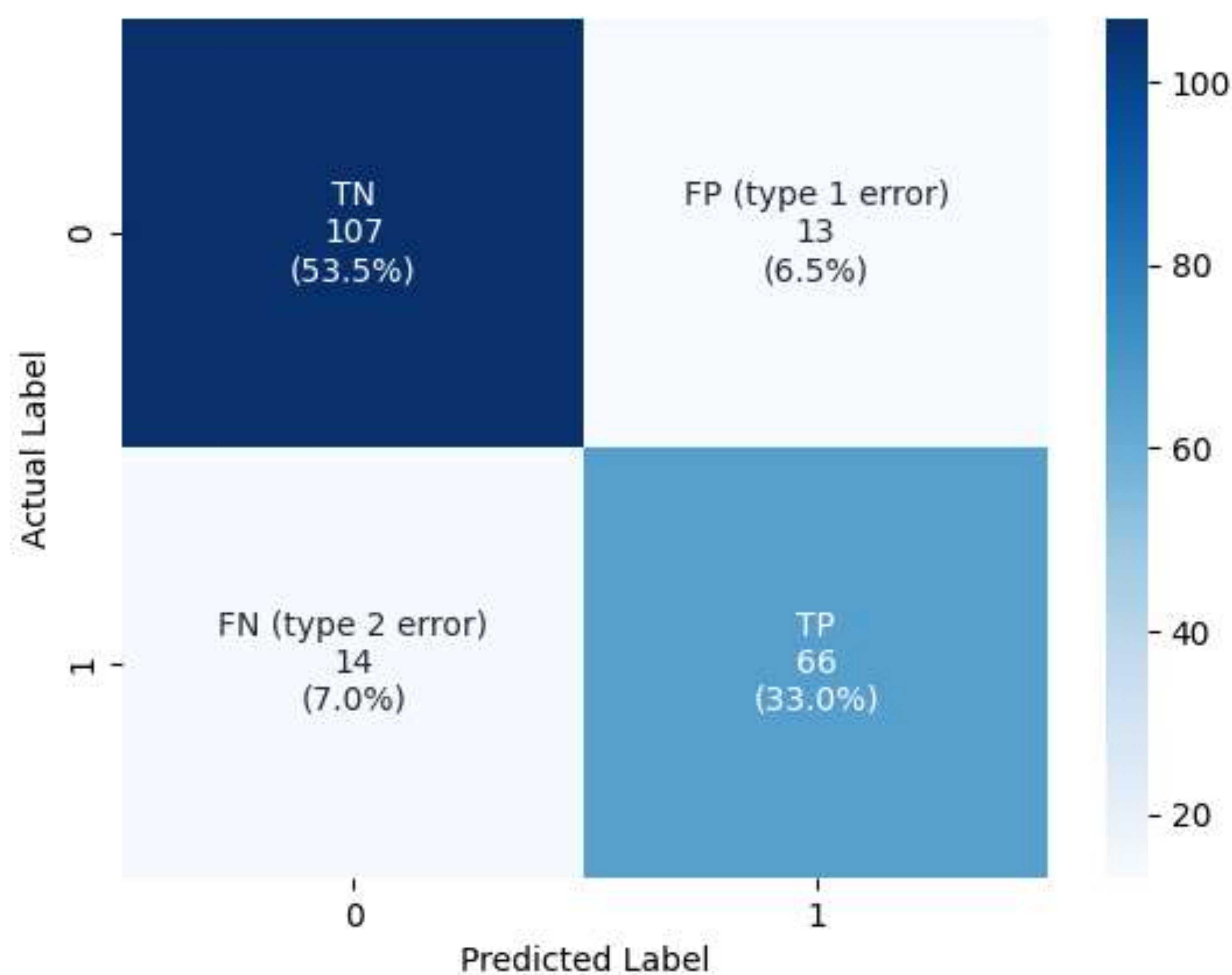
- LPG가스 누출 탐지를 위한 의사결정 나무 머신 러닝 모델과 그 결과에 대한 성능 지표 분석은 100%에 가까운 정확도를 도출. 현재는 1000개의 가상 데이터를 분석하였으나 실제 현장에서는 유해가스탐지기 등을 이용해 실시간으로 정보를 입력 받고 그 정보를 바탕으로 LPG가스 누출을 탐지할 수 있는 프로그램으로 응용 가능할 것으로 전망.





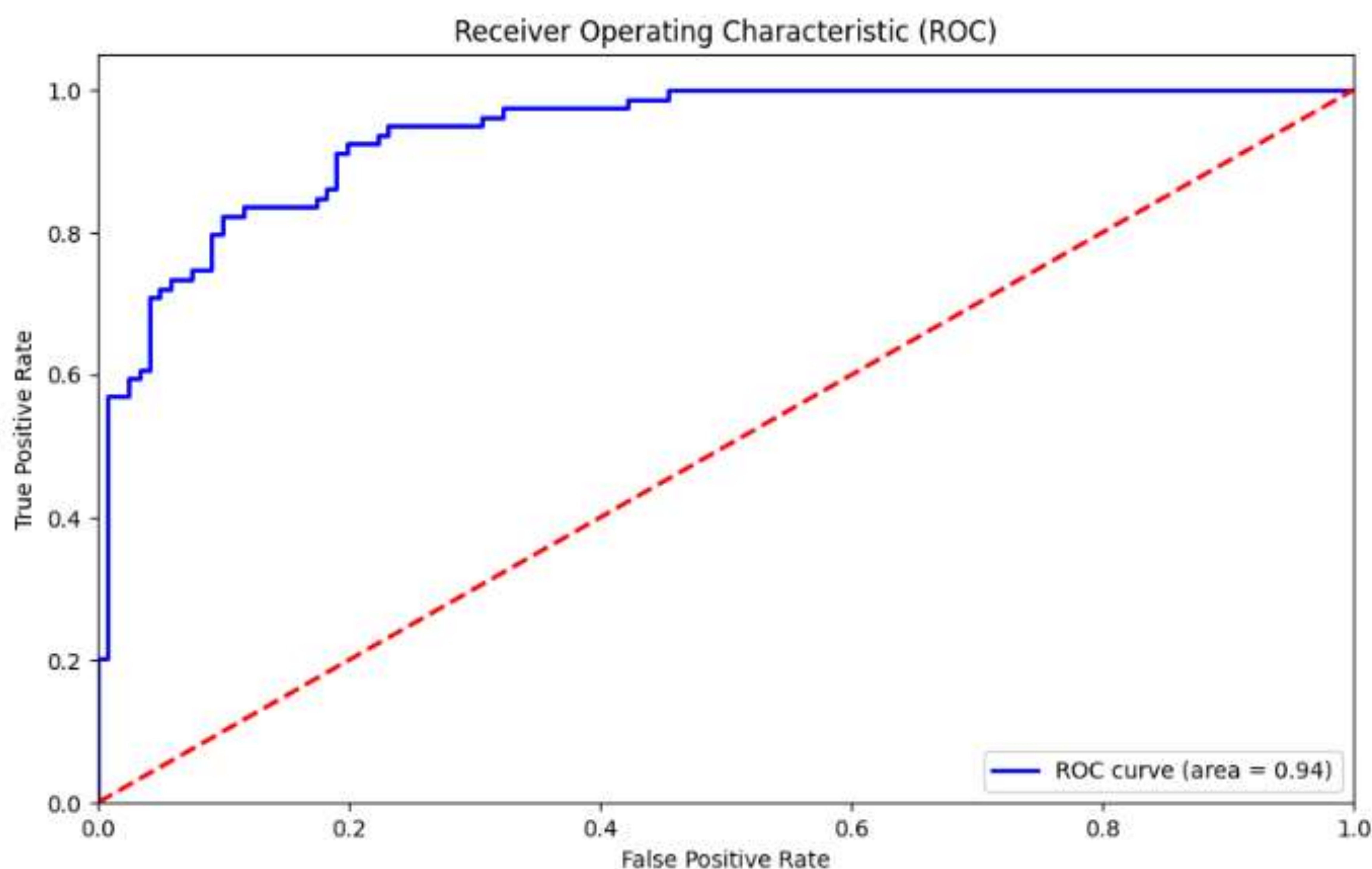
01 머신 러닝 결과

- 이 그래프에서 X축은 8개의 독립 변수로 구성된 1000개의 데이터 중 20%를 분할하여 얻은 200개의 테스트 데이터
- Y축은 LPG 누출 여부를 1(누출)과 0(누출 없음)
- 로지스틱 회귀 모델을 적용한 결과, 실제 값(파란색 원형 점)과 예측 값(빨간색 X형 점)이 86.5%의 정확도로 일치



02 머신 러닝 성능지표 Accuracy 결과

- 이 그래프에서 X축은 예측 값(1:누출, 0:누출 없음)을 Y축은 실제값 (1:누출, 0:누출 없음)
- True Positive(TP) : 실제 누출이 발생한 경우를 정확히 누출로 예측한 비율
- True Negative(TN) : 실제 누출이 발생하지 않은 경우를 정확히 누출 없음으로 예측한 비율
- False Positive(FP) : 실제 누출이 없지만 누출이 있다고 잘못 예측한 비율
- False Negative(FN) : 실제 누출이 발생했지만 누출이 없다고 잘못 예측한 비율



03 머신 러닝 성능지표 ROC_Curve 결과

- 이 그래프에서 X축은 False Positive Rate(FRR)을 나타내며 이는 실제로 부정적인 사례 (누출이 없는 경우)인데 모델이 누출로 잘못 예측한 경우를 의미
- Y축은 True Positive Rate(TPR)을 나타내며 이는 실제로 긍정적인 사례 (누출이 있는 경우)로 모델이 누출로 정확히 예측한 경우를 의미
- ROC 곡선이 계단 모양을 보이고 있으며, 이는 모델의 예측 성능이 일정 구간에서는 변동이 크지 않음을 나타냄, 따라서 모델의 정확도가 비교적 높지 않다고 해석 가능

01 머신 러닝 결과

- 이 그래프에서 X축은 8개의 독립 변수로 구성된 1000개의 데이터 중 20%를 분할하여 얻은 200개의 테스트 데이터
- Y축은 LPG 누출 여부를 1(누출)과 0(누출 없음)
- 로지스틱 회귀 모델을 적용한 결과, 실제 값(파란색 원형 점)과 예측 값(빨간색 X형 점)이 98.5%의 정확도로 일치

02 머신 러닝 성능지표 Accuracy 결과

- 이 그래프에서 X축은 예측 값(1:누출, 0:누출 없음)을 Y축은 실제값 (1:누출, 0:누출 없음)
- True Positive(TP) : 실제 누출이 발생한 경우를 정확히 누출로 예측한 비율
- True Negative(TN) : 실제 누출이 발생하지 않은 경우를 정확히 누출 없음으로 예측한 비율
- False Positive(FP) : 실제 누출이 없지만 누출이 있다고 잘못 예측한 비율
- False Negative(FN) : 실제 누출이 발생했지만 누출이 없다고 잘못 예측한 비율

03 머신 러닝 성능지표 ROC_Curve 결과

- 이 그래프에서 X축은 False Positive Rate(FRR)을 나타내며 이는 실제로 부정적인 사례 (누출이 없는 경우)인데 모델이 누출로 잘못 예측한 경우를 의미
- Y축은 True Positive Rate(TPR)을 나타내며 이는 실제로 긍정적인 사례 (누출이 있는 경우)로 모델이 누출로 정확히 예측한 경우를 의미
- ROC 곡선이 좌상단에 가까워지고 있으며, 이는 모델의 예측 성능이 상대적으로 좋다는 것을 나타냄, 따라서 모델의 정확도가 비교적 높다고 해석 가능

