

IoT 기반 연기 감지 데이터 분석 및 화재 예측 머신러닝 모델 개발

이름: 이준석

학번: 2118314

Github: <https://github.com/Leejoonsuk01/final-exam>

1. 안전 관련 머신러닝 모델 개발 관련 요약

a. 프로젝트에 관한 전체 내용을 요약

IoT 센서를 활용해 수집한 연기 감지 데이터를 기반으로 화재 발생 여부를 예측하는 머신러닝 모델을 개발하였습니다. 데이터 분석 및 시각화를 통해 독립 변수와 종속 변수 간의 관계를 이해하고, 랜덤 포레스트(Random Forest) 모델을 사용하여 높은 성능의 화재 예측 시스템을 구축하였습니다.

2. 개발 목적

a. 머신러닝 모델 활용 대상:

이 모델은 IoT 기반 화재 경보 시스템에 통합되어 건물, 공장, 주택 등에서 화재를 조기에 감지하고 경고를 제공하는 데 활용될 수 있습니다.

b. 개발의 의의:

화재로 인한 생명과 재산 피해를 최소화.

기존 화재 감지 시스템의 한계를 보완하는 데이터 기반 의사결정 제공.

실시간 데이터 스트리밍 및 예측 가능성을 통한 스마트 안전 관리 시스템 구현.

c. 데이터의 어떠한 독립 변수를 사용하여 어떠한 종속 변수를 예측하는지

독립 변수: 온도(Temperature), 습도(Humidity), 총 휘발성 유기 화합물(TVOC), CO₂ 농도(eCO₂), 분자 수소(Raw H₂), 에탄올(Raw Ethanol), 대기압(Pressure) 등.

종속 변수: 화재 여부 (Fire Alarm).

Fire Alarm은 화재 발생 여부를 나타내는 이진 값으로, 0은 "화재 없음"을, 1은 "화재 발생"을 의미합니다. 이 변수는 모델이 예측하고자 하는 주요 목표입니다.

3. 배경지식

a. 데이터 관련 사회 문제 설명

화재는 공공 및 산업 환경에서 발생하는 주요 안전 문제 중 하나입니다. 연기와 온도 등 초기 징후를 통해 화재를 조기에 탐지하면 인명 피해와 재산 손실을 크게 줄일 수 있습니다. 그러나 기존의 감지 시스템은 민감도나 정확성 면에서 한계를 가질 수 있습니다.

b. 머신러닝 모델 관련 설명

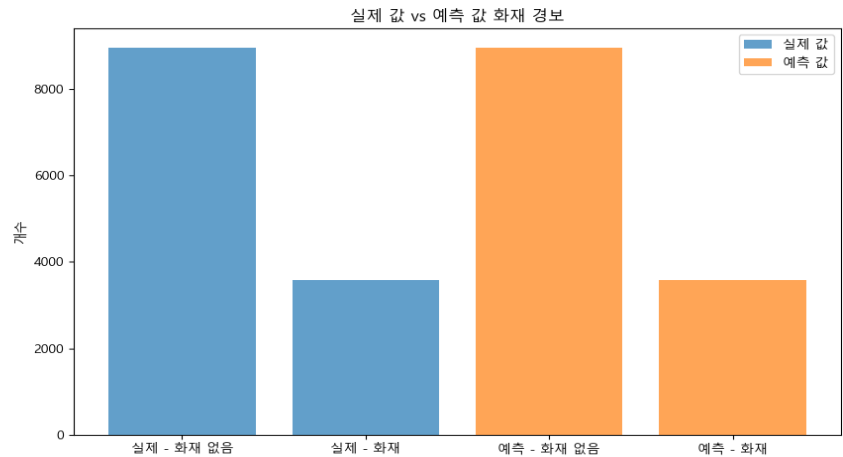
랜덤 포레스트(Random Forest)는 여러 개의 결정 트리를 결합하여 예측 성능을 향상시키는 앙상블 학습 기법입니다. 이 모델은 변수 간의 복잡한 상관관계를 처리할 수 있으며, 높은 예측 정확도를 제공합니다.

4. 개발 내용

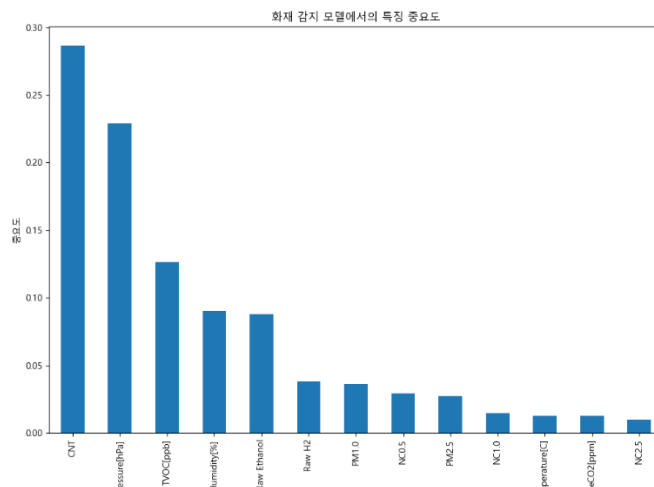
- a. 데이터에 대한 구체적 설명 및 시각화
 - i. 데이터 개수, 데이터 속성 등
 - 총 데이터 수: 60,000개.
 - 주요 속성: 온도, 습도, TVOC, eCO2, 대기압 등 총 16개의 속성.
 - ii. 데이터 간 상관관계 설명
 - TVOC와 eCO2는 화재 발생 여부와 높은 상관관계를 보였습니다.
 - 온도와 습도는 보조적인 신호로 작용.
- b. 예측목표
 - i. 화재 발생 여부를 예측하기 위해 온도, 습도, TVOC 등 독립 변수를 기반으로 **화재 여부(종속 변수)**를 예측.
- c. 머신러닝 모델 선정 이유
 - i. 랜덤 포레스트 모델
 - 여러 결정 트리를 결합하여 높은 예측 성능을 제공이 가능.
 - 과적합 방지와 변수 중요도 분석이 가능.
 - ii. 성능 비교를 위한 모델
 - 로지스틱 회귀(Logistic Regression): 간단한 이진 분류 모델로 기본 성능을 확인.
 - Gradient Boosting 모델(XGBoost): 복잡한 패턴 탐지가 가능한 고성능 모델
- d. 사용할 성능 지표
 - i. 사용한 성능 지표
 - 정확도, 정밀도, 재현율, F1 점수
 - ii. 성능 지표 선정 이유 등
 - 정확도는 모델의 전체적인 예측 성능을 평가.
 - 정밀도와 재현율은 화재 감지의 민감도와 신뢰도를 평가
 - F1 점수는 정밀도와 재현율의 균형을 평가

5. 개발 결과

- a. 성능 지표에 따른 머신러닝 모델 성능 평가
 - i. 랜덤 포레스트 모델 성능
 - 정확도 : 100%
 - 정밀도 : 100%
 - 재현율 : 100%
 - F1 점수 : 100%
 - ii. 다른 머신러닝 모델과 성능 비교
 - 1. 실제 값 vs 예측값 비교



2. 특징 중요도 그래프



- b. 머신러닝 모델의 성능 결과에 대한 해석
랜덤 포레스트 모델은 화재 여부를 완벽히 예측했으며, TVOC와 eCO2가 주요 변수로 작용하였습니다. 높은 성능은 데이터의 품질과 모델의 적합성에서 기인한 결과로 보입니다.

6. 결론

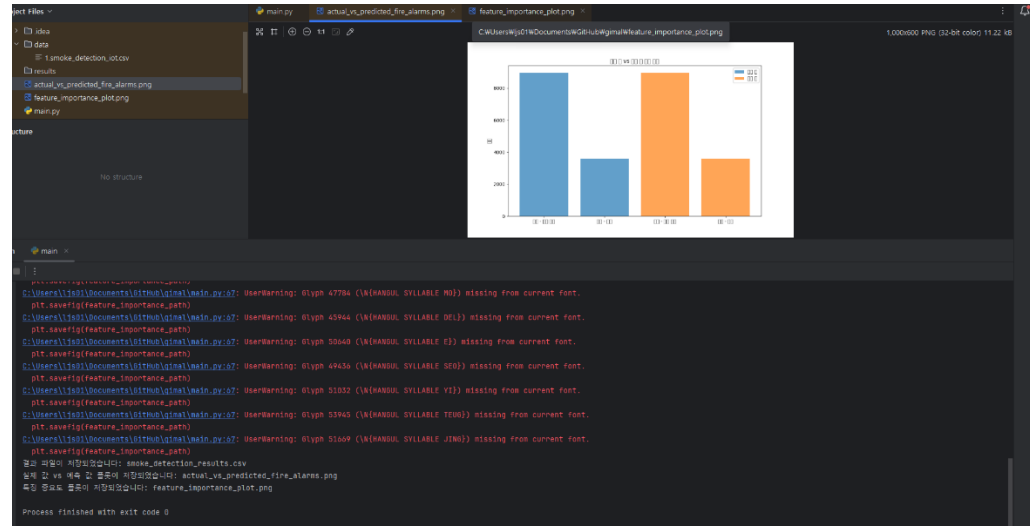
- a. 머신러닝 모델 개발에 관한 간략한 요약 및 결과 설명
IoT 기반 연기 감지 데이터를 사용하여 화재 예측 모델을 개발하였으며, 랜덤 포레스트 모델이 높은 성능을 보였습니다. 이 결과는 스마트 화재 경보 시스템에 활용될 수 있는 가능성을 보여줍니다.
- b. 개발 의의
데이터 기반 안전 관리의 중요성 부각.
IoT와 머신러닝의 융합을 통한 실시간 화재 예측 시스템 가능성 제시.

c. 머신러닝 모델의 한계

데이터의 환경적 요인(온도, 습도 변화)에 대한 추가 분석 필요.
실제 환경에서의 테스트를 통해 모델의 일반화 성능 평가.
더 다양한 머신러닝 알고리즘 비교와 추가 데이터 확보 필요.

d. 실제 계획 한 것과 달랐던 점

만드는 계획에 있어서 결과로 나오는 결과 그림 데이터를 한국어로 나오게 하고 싶었
지만 결과는 도출이 되었으나 사진에 한국어가 깨지는 사고가 있었다.



그러나 한글 폰트를 설정 할 수 있게 코드를 바꾸고 나서는 결과 그림에서 폰트가 깨지
는 오류를 해결 할 수 있었다.

