Fire Prediction Model

**학번: 2018005**

**이름: 김민성**

**Github address: https://github.com/minsungkim333333/Fire-Prediction-Model**

1. **안전 관련 머신러닝 모델 개발의 목적** 
   1. 화재유형별 피해를 분석하여 화재 예방에 도움을 준다.
   2. 학습 모델 활용 대상 : 소방처 등
   3. 사망, 부상자, 재산피해, 인명피해, 발생 시간을 독립변수로 하고 화재 유형을 종속변수로 하였다.
   4. 개발의 의의: 화재 발생을 예측하여 피해를 입기 전에 예방할 수 있다.
2. **안전 관련 머신러닝 모델의 네이밍의 의미** 
   1. Fire Prediction Model은 화재 예측 모델의 영문명으로 사망, 부상자, 재산피해, 인명피해, 발생 시간 등을 이용하여 화재유형을 예측한다. 직관적인 이름이 좋을 것 같아 화재 예측 모델이라고 네이밍을 하였다.
3. **개발 계획**
   1. 데이터 전처리 계획
      1. 데이터의 정보를 확인하였다.텍스트, 스크린샷, 메뉴이(가) 표시된 사진

         자동 생성된 설명
      2. 화재 유형별 일시, 인명피해(명)소계 등 정수형 자료들을 사용한다.
   2. 머신러닝 모델 : 화재 유형별 수치들을 분류하는 작업이기 때문에 **RandomForestClassifier** 라는 분류 작업에 특화된 모델을 사용하였다. 이 모델은 의사결정 트리(Decision Tree)를 기반으로 하며, 여러 개의 의사결정 트리를 조합하여 높은 정확도와 일반화 성능을 제공한다.
      1. 앙상블 모델: 여러 개의 결정 트리를 함께 사용하는 앙상블 모델이다. 이러한 결정 트리들을 랜덤한 서브셋으로 학습하고 그 결과를 종합하여 예측을 수행한다.
      2. 부트스트랩 샘플링: 각 트리는 원본 데이터셋에서 랜덤하게 선택된 부트스트랩 샘플(복원추출)을 사용하여 학습한다. 이로써 각 트리는 서로 다른 관측치들을 학습하게 된다.
      3. 랜덤 특성 선택: 각 트리는 랜덤하게 선택된 특성들을 사용하여 분할을 수행한다. 이는 특성의 다양성을 증가시켜 모델이 더욱 강건하게 되도록 도와준다.
      4. 다양한 의사결정 트리의 결합: 각 트리가 독립적으로 학습하고 예측하므로, 다양한 관점에서 데이터를 해석하고 학습할 수 있다. 이를 통해 과적합(Overfitting)을 줄이고 일반화 성능을 향상시킨다.
      5. Out-of-Bag 평가: 부트스트랩 샘플링을 통해 생성되지 않은 샘플들을 사용하여 각 트리의 성능을 평가하는데 사용됩니다. 이를 통해 교차 검증을 수행하지 않고도 모델의 성능을 추정할 수 있다.
   3. 사용할 성능 지표
      1. Accuracy, precision, recall, f1-score을 이용하여 성능을 확인하려고 한다.
   4. 성능 검증 방법 계획 등

K-fold 교차검증을 사용하여 성능을 확인할 예정이다.

1. **개발 과정** 
   1. **화재 유형의 종류 확인**

텍스트, 폰트, 스크린샷이(가) 표시된 사진

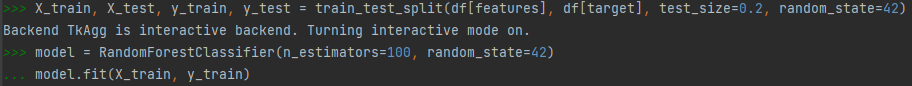
자동 생성된 설명

****

**독립, 종속 변수 설정**

****

**데이터 분할, 모델 선택 및 학습**

****

**모델 평가**

**텍스트, 스크린샷, 소프트웨어이(가) 표시된 사진

자동 생성된 설명**

**혼동 행렬 계산 및 시각화**

**텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 스크린샷, 번호, 평행이(가) 표시된 사진

자동 생성된 설명**

**혼동행렬 항목들이 박스로 뜨는 오류 발생**

****

**한글 폰트가 없다고 판단, 가능한 한글 폰트 적용**

****

**텍스트, 스크린샷, 도표, 평행이(가) 표시된 사진

자동 생성된 설명**

**오류 해결**

**클래스 별 정밀도,재현율 f1 계산 후 시각화**

**텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명**

**텍스트, 스크린샷, 도표, 그래프이(가) 표시된 사진

자동 생성된 설명**

* 1. **K-fold 교차 검증텍스트, 스크린샷, 폰트, 소프트웨어이(가) 표시된 사진

     자동 생성된 설명텍스트, 스크린샷이(가) 표시된 사진

     자동 생성된 설명**
  2. **결과**
     1. 전체적으로 정확도가 높지 않아, 더 많은 자료나 다른 모델을 사용하여 정확도를 개선할 필요가 있다고 생각된다.
     2. 결과를 보면 건축, 구조물에서 높은 정확도를 보여준다. 그렇기 때문에 이 모델은 건축 구조물을 대상으로 활용할 수 있다고 생각한다.
     3. 건축, 구조물에서 화재 원인을 예측하는 모델로서 활용이 가능하다.

1. **개발 후기**

데이터를 조사하는 과정에서부터 어려움을 느꼈다. 활용성이 높은 데이터, 데이터 간의 연관성이 높은 데이터들의 가치가 높은 이유를 알게 되었다.

데이터 사이의 연관성을 파악하는 과정들이 생각보다 흥미롭고 즐거웠다.

시각화를 통해서 직관적으로 확인할 수 있어, 미래에 데이터를 분석하는 기회가 온다면 잘 활용할 수 있을 것이라고 생각했다.