산불 발생단계 분석

**먼저 04\_허건호.csv 파일에서 분석에 필요한 feature만 남김**

분석에 사용한 feature

**화재발생\_월, 주말여부, 화재발생시간대, 시군구명, 발화지점, 날씨, 온도 습도, 발화원인 종합**

**피해면적을 예측해야 하므로 모델은 DecisionTreeRegressor 사용함**

DecisionTreeRegressor은 숫자형 데이터를 입력으로 받기 때문에 문자열 데이터들을 ONE-hot encoding으로 변환함

**Datasplit: ( train, validation, test = 0.63:0.27:0.1) 로 split 함**

**아래 3가지 hyperparameter를 가지고 GridSearchCV 를 통한 적절한 hyperparameter 찾아보았지만**

1. max\_depth: 의사결정나무의 최대 깊이를 지정하는 파라미터입니다. 깊이가 너무 깊으면 모델이 train 데이터에 overfitting 될 가능성이 높아집니다. 적절한 값을 찾아주어야 합니다.

2. min\_samples\_split: 노드를 분할하기 위해 필요한 최소 샘플 수를 지정하는 파라미터입니다. 이 값을 높게 설정하면 모델이 train 데이터에 overfitting 되는 것을 방지할 수 있습니다.

3. min\_samples\_leaf: 말단 노드(leaf)가 되기 위한 최소한의 샘플 수를 지정하는 파라미터입니다. 이 값을 높게 설정하면 모델이 train 데이터에 overfitting 되는 것을 방지할 수 있습니다.

유의미한 결과는 나오지 않았습니다

코드에 나온 수치들 외에도 다양하게 시도해보았으나 계속해서 높은 error 값이 나오네요.

차원의 저주 땜시인가?, 앙상블 시도? 변수 변화?, 차원 축소?