

<https://dacon.io/competitions/official/236038/codeshare/7304?page=1&dtype=recent>

데이터

- train
 - y_var : 착과량 (예측 목표 변수)
 - saesoon: 새순 (시계열 데이터 변수)
 - yuprokso : 엽록소 (시계열 데이터 변수)
- test

코드 흐름

- (1) 라이브러리 불러오기
: pandas, numpy, MinMaxScaler, train_test_split, RandomForestRegressor
- (2) 데이터 로드 및 변수 설정
 - y_var : 예측할 목표 변수인 착과량(int)
 - id_var, saesoon, yuprokso : ID, 새순(시계열 데이터 변수), 엽록소(시계열 데이터 변수)
- (3) 전처리
 - 데이터 멜트(melt) : melt함수를 이용해 새순과 엽록소 데이터를 단일 열로 변환하여 date와 type로 구분. 시계열 데이터를 각 날짜별로 처리할 수 있도록 함.
 - 날짜와 유형 열 생성: variable열의 값을 date와 type로 분리해서 각각 새로운 열로 추가함.
 - 전처리 데이터 셋 생성: 새순diff0, 새순diff1, 엽록소_새순 같은 새로운 피처를 생성.
- (4) 피벗 테이블 변환
: pivot을 사용해 각 날짜별로 데이터를 정리하여 학습/테스트 데이터 셋 생성.
- (5) 데이터 학습 및 예측
: RandomForestRegressor 모델을 학습시킴(max_depth=5로 설정해서 복잡도 조절).

인상깊었던점 및 배울점:

시계열 데이터의 전처리와 피처 엔지니어링 과정이 특히 체계적이다.

- melt 함수를 사용해 시계열 데이터를 '넓은' 형식에서 '긴' 형식으로 변환 후, 각 날짜별 데이터를 pivot으로 재구성.
- 엽록소/새순, 새순i-새순(i+1)처럼 시계열적 변화에 적합한 새로운 변수를 생성한 점이 인상적이다.