

[청경채 성장 예측 AI경진대회]

[청경채 성장 예측 AI 경진대회 - DACON](#)
참고 코드: [\[private 2nd\] 테라파머 - DACON](#)

❖ 데이터

● train_input:

총 58개 청경채 케이

각 청경채 케이스 별 환경 데이터(1분 간격)

시간(1분간격), 내부온도관측치, 내부습도 관측치, CO2 관측치, EC 관측치, 외부온도 관측치, 외부습도 관측치, 펌프상태, 펌프 작동 남은시간, 최근분무량, 일간 누적분무량, 냉방상태, 냉방작동 남은시간, 난방상태, 난방작동남은시간, 내부유동팬상태, 내부유동팬작동남은시간, 외부환기팬상태, 외부환기팬작동남은시간, 화이트 LCD상태, 화이트 LED작동남은시간, 화이트 LED동작강도, 레드 LED 상태, 레드LED 작동남은시간, 레드 LED동작강도, 블루 LED상태, 블루 LED작동남은시간, 블루 LED동작강도, 카메라상태, 냉방온도, 난방온도, 기준온도, 난방부하, 냉방부하, 총추정광량, 백생광추정광량, 적색광추정광량, 청생광추정광량

● train_target:

총 58개 청경채 케이스

rate : 각 청경채 케이스 별 잎 면적 증감률(1일 간격)

● test_input:

총 6개 청경채 케이스

각 청경채 케이스 별 환경 데이터(1분 간격)

● test_target:

총 6개 청경채 케이스

rate: 각 청경채 케이스 별 잎 면적 증감률(1일간격)

❖ 코드흐름

1. 패키지 임포트
2. 전처리
 - 1) 분단위 데이터를 60분씩 평균해서 24시간 데이터 만들기
 - 2) 시간단위 데이터에서 일단위 합계 데이터 추가
 - 3) 시간단위 데이터를 일단위 데이터로 변환

- test_input은 case별로 데이터 사용, train_input은 58개 합쳐서 하나로 사용

3. 탐색적 자료분석 EDA (Exploratory Data Analysis) 및 변수선택
4. 하루 1440분(60*24)데이터 살펴보기 -> 학습에 영향을 줄 피쳐 확인
5. Target 변수 살펴보기 (이상치 제거)
6. 모델 구축, 학습
 - lightGBM 모델
 - train, validation, testing 세트로 이루어진 학습데이터를 읽고 하이퍼파라미터를 변경하면서 rmse계산, rmse가 가장 적은 파라미터 반환

7. 학습결과 확인
 - 상위 10개 피처 중요도, **metric** 측정결과, 실측치와 예측치의 상관도
8. 결론
 - 타겟변수에 직접적인 영향을 주는 변수만 선택하고 데이터를 60분 단위로 만든 후 LGBM으로 학습해 예측하는 것이 가장 성적에 좋음
9. 추가제언

❖ 배운점

- 피처 목록을 통해 재배시절을 예상하고 변수 조절을 할 수 있다.
- 데이터양이 일정수준 이상으로 있는 피처들만 사용함(막대그래프로 확인)
- 스캐터 플랏을 그려 이상치를 확인, **lineplot**으로 변수변화 확인
- 시간 데이터를 다룰 때 30분, 15분, 5분 같이 여러 단위로 변환해 학습하면 성적이 다를 수 있다.
- 최적의 학습을 이룰 수 있도록 필요한 변수만 선택하는 것이 중요한 것 같다.
- 무조건 기법을 추가하고 많은 피처를 선택할수록 최적의 결과가 나오는 것은 아니다.
- 하이퍼 파라미터 튜닝에 **rmse**를 활용할 수 있다.