

# 2023-05-16 업무분담

🕒 생성일	@2023년 5월 16일 오후 4:33
🏷 태그	

## 1. 곡물 생산량 예측 모델 제작

- 일단 random forest로 가중치를 구해서 3개 Feature에 대해서 모델 제작
  - 일조량, 채소소비량, 육류소비량, 비료 가격, 유가
- 추후 2개 추가할 수 있음.

## 2. 어떤 Feature가 중요한지 파악 → 헤림님

1. data 가지고 random forest를 사용하여 가중치 비교
2. EDA 도 진행

## 3. 다른 Feature 데이터에 대해 예측 모델 제작

1. 기후라면 LSTM
2. 다른 데이터라면 무엇?

## 4. 추가 데이터 수집 or Data Augmentation → 동준님

1. 곡물 생산량에 대한 데이터 추가 제작
2. 작물 가격, Co2 데이터 포함하여 augmentation 까지 진행



시계열 데이터에 데이터 증강(Data Augmentation)을 적용하는 방법은 다양하지만, 다음과 같은 몇 가지 일반적인 기법들이 있습니다:



1. 시프트(Shift): 시계열 데이터를 시간 축을 따라 일정한 시간 단위로 이동시킵니다. 이를 통해 데이터의 시간적인 위치를 변경할 수 있습니다. 예를 들어, 이전 시점의 데이터를 현재 시점으로 이동시키거나, 현재 시점의 데이터를 미래로 이동시킬 수 있습니다.
2. 스케일링(Scaling): 시계열 데이터의 값을 일정한 비율로 조정합니다. 일반적으로 최소-최대 스케일링이나 표준화(standardization)를 사용합니다. 이를 통해 데이터의 크기나 분포를 조정할 수 있습니다.
3. 회전(Rotation): 시계열 데이터의 주기성을 나타내는 특성을 반영하여 데이터를 회전시킵니다. 회전은 주파수 도메인 변환 기법 중 하나인 FFT(Fast Fourier Transform)를 사용하여 수행할 수 있습니다.
4. 추가적인 데이터 인접화(Additional Data Augmentation): 데이터를 보강하기 위해 다른 관련 데이터를 추가할 수 있습니다. 예를 들어, 이전 기간의 데이터를 현재 기간과 함께 사용하거나, 다른 관련 변수의 데이터를 포함시킬 수 있습니다.
5. 잡음 추가(Adding Noise): 데이터에 잡음을 추가하여 모델의 강건성(robustness)을 향상시킬 수 있습니다. 잡음 추가는 데이터의 변동성을 증가시키고, 모델이 더 일반화된 패턴을 학습할 수 있도록 돕습니다.
6. 삭제(Data Trimming): 장기간 동안 데이터가 없는 구간이나 노이즈가 심한 부분을 제거하여 데이터의 품질을 향상시킵니다. 이는 모델이 잘못된 정보를 학습하지 않도록 돕습니다.

위의 기법들은 시계열 데이터의 특성에 따라 유연하게 조정될 수 있습니다. 데이터 증강 기법을 적용할 때는 주어진 문제와 데이터에 맞게 선택하고 적용해야 합니다. 또한, 증강된 데이터가 원래 데이터와 동일한 통계적 특성을 가지도록 유의해야 합니다.