

커피 가격 예측 모델별 결과 정리

모델별 단계에 대해서는 이전 PDF에 정리되어 있으므로 결과 위주로 정리하였다.

1. VAR

a) 거시경제데이터 이용

- Train 데이터의 마지막 date기준으로 2주후 후까지를 예측



train 데이터의 마지막 시점을 기준으로 2주후를 예측하여 실제 값과 비교해보았다.

- 모델 평가

모델 성능평가 지표로 MAE와 RMSE를 사용한다.

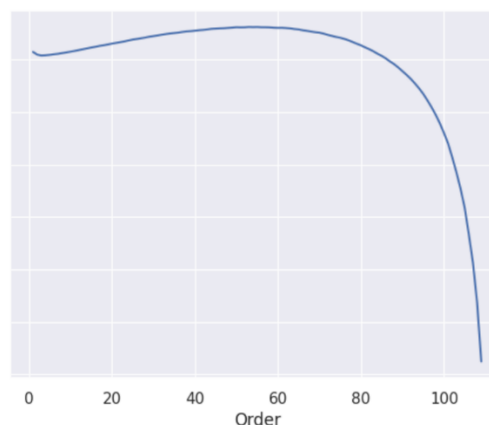
MAE	RMSE
5.48	6.32

가격 수준이 200 정도라면, 약 2-3% 정도 오차가 나타난다.

b) 기후데이터 사용

VAR 모델은 모든 시계열 변수에 대해 자체적으로 lag구조를 내부적으로 생성한다. 따라서 lag 피처를 생성하지 않은 원본데이터에 대해서만 학습을 진행한다.

- 최적의 lag값 파악을 위한 lag값에 따른 AIC 변화 그래프



기후는 커피 생산량에 영향을 주는데 실제 가격에 반영되기까지는 수개월이 걸릴 수 있다. 따라서 기후 변수가 단기 예측에 거시경제지표에 비해 덜 직접적일 수 있다.

기후는 커피 가격에 즉각적인 영향을 주지 않고, 수개월 혹은 수십주 뒤에 영향을 주는 경우가 많다. 따라서 모델은 긴 시간 범위의 정보를 활용해야 더 정확하다고 판단하므로 VAR은 긴 lag값을 선택하려고 한다.

하지만 VAR에서 lag가 많아질수록 모수가 급격히 증가해 과적합의 가능성이 있다.

따라서 선형 모델이 아닌 비선형 모델을 이용하거나 **생산량 지수 피처를 만드는 방식을 활용**하는게 더 현실적일 수 있다.

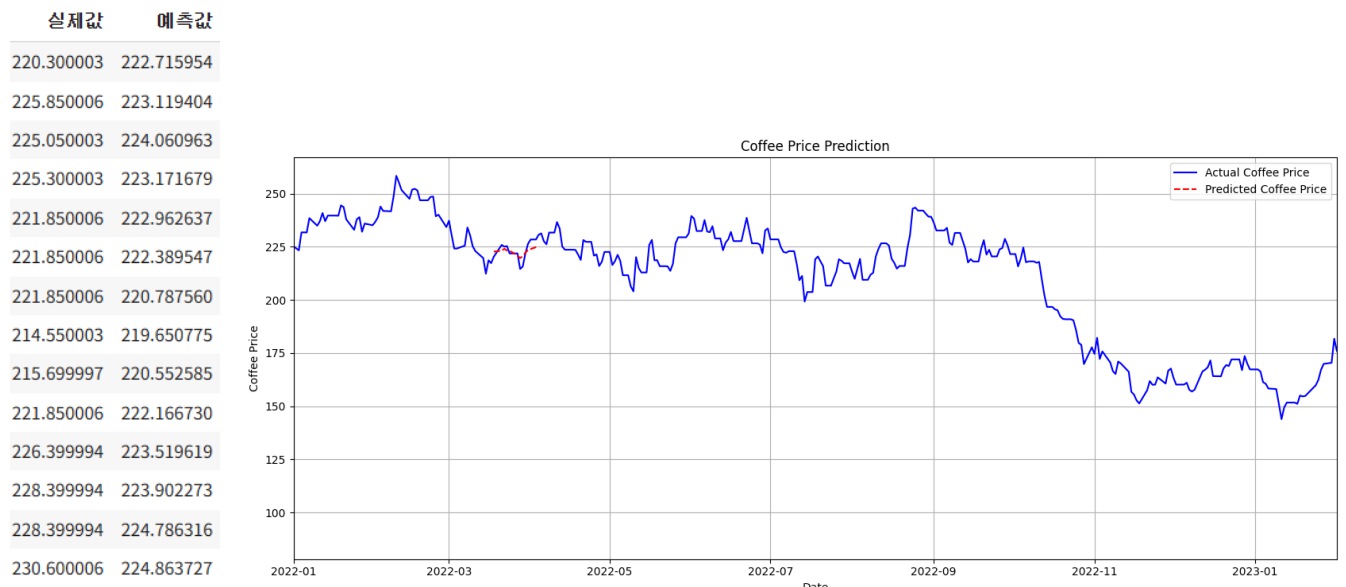
기후데이터를 직접 가격에 영향을 주기 전에 생산량에 영향을 준다. 기후 -> 생산량 -> 가격의 흐름이 현실적이다. 기후 데이터를 통해 커피 생산량 추정 지수를 만들어주고 이를 커피 가격 예측을 위한 변수로 사용하여 예측을 진행하는 방식을 이용할 수 있다.

2. LSTM

a) 거시경제데이터 이용

데이터셋을 구성할 때 최근 180개 데이터를 입력으로 사용하였다.

- Train 데이터의 마지막 date기준으로 2주후 후까지를 예측



- 모델 평가

MAE	RMSE
2.71	3.25

가격 수준이 200 정도라면, 약 1% 정도 오차가 나타난다.

b) 거시경제+기후데이터

1. 기후데이터 원본 그대로 사용

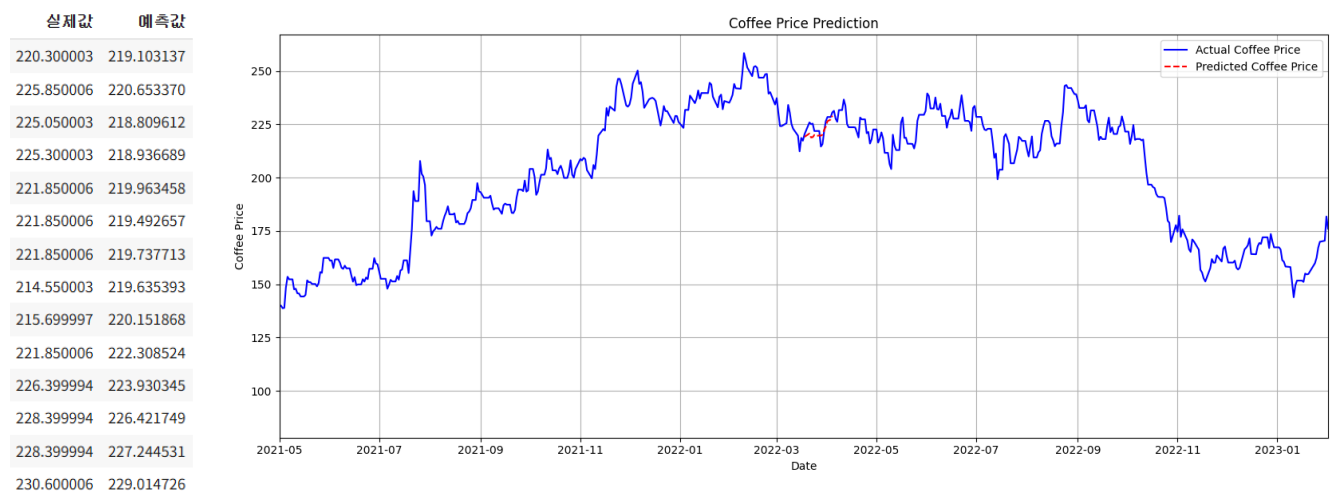
기존 거시경제데이터에 기후데이터를 추가하였을 때 기후데이터가 예측에 유의미한 정보를 추가적으로 주는지에 대해 확인해보았다.

LSTM은 시계열의 순서를 보존하고 시간에 따라 변하는 패턴을 학습하도록 설계되기 때문에 각 시점별 입력을 따로 정렬된 형태로 넣는게 이상적이다.

즉 lag피처를 사용하여 한 input 안에서 피처간 시점이 다르게 하는 것이 아니라 현재 사용하는 방식처럼 하나의 시퀀스 입력으로 학습하고 LSTM이 시계열 구조 속에서 학습하였다.

```
data_window = 180 # 최근 180개 데이터를 입력으로 사용
future_target = 14 # 14개 미래 값 예측
step = 6 # 6개 단위로 샘플링
train_dataset = MultiStepTimeSeriesDataset(X_train, y_train, data_window, future_target, step)
```

- Train 데이터의 마지막 date기준으로 2주후 후까지를 예측



- 모델 평가

MAE	RMSE
3.03	3.59

거시경제데이터만 넣었을 때보다 후반부의 정확도는 높지만 초반에 실제값과의 차이가 있어 거시경제로만 했을 때 보다 MAE와 RMSE가 0.3정도 낮게 나온다.

기후데이터를 사용하는 것이 예측값에 어떠한 중요도를 가지는지를 통합 그래디언트를 통해 알아보았다.

통합그래디언트는 입력값이 바뀔 때 예측이 얼마나 달라지는지 경사를 누적하여 변수 중요도를 평가하는 방법이다.

1일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002729
- brazil_patrocinio_RH2M → contribution: 0.002543
- Coffee_Price → contribution: 0.002308
- Urea_price → contribution: 0.002199
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002044
- Colombia_GDP per capita growth (annual %) → contribution: 0.002042
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.001998
- USD_BRL → contribution: 0.001599
- brazil_varginha_RH2M → contribution: 0.001584
- brazil_carmo_de_minas_RH2M → contribution: 0.001565

2일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002789
- brazil_patrocinio_RH2M → contribution: 0.002484
- Coffee_Price → contribution: 0.002339
- Urea_price → contribution: 0.002136
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002083
- Colombia_GDP per capita growth (annual %) → contribution: 0.002063
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002055
- USD_BRL → contribution: 0.001543
- Brazil_GDP per capita growth (annual %) → contribution: 0.001543
- brazil_carmo_de_minas_RH2M → contribution: 0.001542

3일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002747
- brazil_patrocinio_RH2M → contribution: 0.002345
- Coffee_Price → contribution: 0.002321
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002048
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002046
- Colombia_GDP per capita growth (annual %) → contribution: 0.002012
- Urea_price → contribution: 0.001984
- Brazil_GDP per capita growth (annual %) → contribution: 0.001525
- USD_BRL → contribution: 0.001459
- brazil_carmo_de_minas_RH2M → contribution: 0.001455

7일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002811
- brazil_patrocinio_RH2M → contribution: 0.002536
- Coffee_Price → contribution: 0.002456
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002190
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002082
- Colombia_GDP per capita growth (annual %) → contribution: 0.001999
- Urea_price → contribution: 0.001982
- Brazil_GDP per capita growth (annual %) → contribution: 0.001588
- USD_BRL → contribution: 0.001579
- brazil_varginha_RH2M → contribution: 0.001555

8일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002811
- brazil_patrocinio_RH2M → contribution: 0.002526
- Coffee_Price → contribution: 0.002462
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002206
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002069
- Colombia_GDP per capita growth (annual %) → contribution: 0.001987
- Urea_price → contribution: 0.001978
- USD_BRL → contribution: 0.001609
- Brazil_GDP per capita growth (annual %) → contribution: 0.001590
- brazil_varginha_RH2M → contribution: 0.001582

9일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002824
- brazil_patrocinio_RH2M → contribution: 0.002631
- Coffee_Price → contribution: 0.002477
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002227
- Urea_price → contribution: 0.002112
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002069
- Colombia_GDP per capita growth (annual %) → contribution: 0.001987
- USD_BRL → contribution: 0.001695
- brazil_varginha_RH2M → contribution: 0.001655
- Brazil_GDP per capita growth (annual %) → contribution: 0.001595

13일 후 예측값에 영향을 준 Top 10 Features:

- brazil_patrocinio_RH2M → contribution: 0.003238
- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002987
- Coffee_Price → contribution: 0.002570
- Urea_price → contribution: 0.002510
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002380
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002137
- Colombia_GDP per capita growth (annual %) → contribution: 0.002103
- brazil_varginha_RH2M → contribution: 0.002014
- USD_BRL → contribution: 0.001937
- brazil_carmo_de_minas_RH2M → contribution: 0.001882

14일 후 예측값에 영향을 준 Top 10 Features:

- brazil_patrocinio_RH2M → contribution: 0.003420
- Colombia_Food production index (2014-2016 = 100) → contribution: 0.003036
- Urea_price → contribution: 0.002596
- Coffee_Price → contribution: 0.002580
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002419
- brazil_varginha_RH2M → contribution: 0.002160
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002157
- Colombia_GDP per capita growth (annual %) → contribution: 0.002145
- USD_BRL → contribution: 0.002014
- brazil_carmo_de_minas_RH2M → contribution: 0.001984

4일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002770
- brazil_patrocinio_RH2M → contribution: 0.002431
- Coffee_Price → contribution: 0.002356
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002092
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002063
- Colombia_GDP per capita growth (annual %) → contribution: 0.002010
- Urea_price → contribution: 0.001954
- Brazil_GDP per capita growth (annual %) → contribution: 0.001546
- USD_BRL → contribution: 0.001511
- brazil_varginha_RH2M → contribution: 0.001467

5일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002815
- brazil_patrocinio_RH2M → contribution: 0.002448
- Coffee_Price → contribution: 0.002402
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002151
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002087
- Colombia_GDP per capita growth (annual %) → contribution: 0.002025
- Urea_price → contribution: 0.001963
- Brazil_GDP per capita growth (annual %) → contribution: 0.001574
- USD_BRL → contribution: 0.001525
- brazil_varginha_RH2M → contribution: 0.001488

6일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002811
- brazil_patrocinio_RH2M → contribution: 0.002510
- Coffee_Price → contribution: 0.002418
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002168
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002078
- Colombia_GDP per capita growth (annual %) → contribution: 0.002007
- Urea_price → contribution: 0.001966
- Brazil_GDP per capita growth (annual %) → contribution: 0.001575
- USD_BRL → contribution: 0.001536
- brazil_varginha_RH2M → contribution: 0.001510

10일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002895
- brazil_patrocinio_RH2M → contribution: 0.002789
- Coffee_Price → contribution: 0.002538
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002293
- Urea_price → contribution: 0.002253
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002110
- Colombia_GDP per capita growth (annual %) → contribution: 0.002032
- USD_BRL → contribution: 0.001762
- brazil_varginha_RH2M → contribution: 0.001737
- Brazil_GDP per capita growth (annual %) → contribution: 0.001632

11일 후 예측값에 영향을 준 Top 10 Features:

- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002907
- brazil_patrocinio_RH2M → contribution: 0.002865
- Coffee_Price → contribution: 0.002524
- Urea_price → contribution: 0.002339
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002308
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002111
- Colombia_GDP per capita growth (annual %) → contribution: 0.002044
- brazil_varginha_RH2M → contribution: 0.001791
- USD_BRL → contribution: 0.001745
- brazil_carmo_de_minas_RH2M → contribution: 0.001643

12일 후 예측값에 영향을 준 Top 10 Features:

- brazil_patrocinio_RH2M → contribution: 0.003006
- Colombia_Food production index (2014-2016 = 100) → contribution: 0.002955
- Coffee_Price → contribution: 0.002545
- Urea_price → contribution: 0.002410
- Ethiopia_Export unit value index (2015 = 100) → contribution: 0.002346
- Colombia_IMF repurchases and charges (TDS, current US\$) → contribution: 0.002131
- Colombia_GDP per capita growth (annual %) → contribution: 0.002074
- brazil_varginha_RH2M → contribution: 0.001882
- USD_BRL → contribution: 0.001825
- brazil_carmo_de_minas_RH2M → contribution: 0.001748

기후데이터를 추가했을 때 기후데이터의 피쳐 중 예측에 기여를 주는 피쳐가 있는지를 확인해보았다.

brazil_patrocinio_RH2M, brazil_varginha_RH2M, brazil_carmo_de_minas_RH2M 같은 기후 변수들 이 꾸준히 상위권에 있다. 이들은 다른 주요 경제 변수들과 비슷하거나 높은 수준으로 예측값에 영향을 보여주고 있다. 즉, 기후변수의 계절성 및 지역성으로 오히려 모델의 예측 불안정을 줄 수 있다.

**** 기후데이터의 피처별 최적의 lag값 확인 및 target변수와의 상관관계 출력****

변수별 최적 Lag 및 상관계수:

	macro_variable	lag	corr			
0	brazil_carmo_de_minas_ALLSKY_SFC_SW_DWN	0	0.033993	17	colombia_manizales_PRECTOTCORR	1 0.061065
1	brazil_carmo_de_minas_PRECTOTCORR	3	-0.051961	18	colombia_manizales_T2M	6 0.228184
2	brazil_carmo_de_minas_T2M	0	0.061238	19	colombia_manizales_WS2M	6 -0.040381
3	brazil_carmo_de_minas_WS2M	2	0.046045	20	colombia_pereira_ALLSKY_SFC_SW_DWN	3 0.050835
4	brazil_patrocinio_ALLSKY_SFC_SW_DWN	3	0.043626	21	colombia_pereira_PRECTOTCORR	1 0.061065
5	brazil_patrocinio_PRECTOTCORR	3	-0.067627	22	colombia_pereira_T2M	6 0.228184
6	brazil_patrocinio_T2M	2	0.123347	23	colombia_pereira_WS2M	6 -0.040381
7	brazil_patrocinio_WS2M	5	-0.103345	24	ethiopia_limu_ALLSKY_SFC_SW_DWN	4 -0.072826
8	brazil_varginha_ALLSKY_SFC_SW_DWN	3	0.030025	25	ethiopia_limu_PRECTOTCORR	3 0.059729
9	brazil_varginha_PRECTOTCORR	3	-0.024711	26	ethiopia_limu_T2M	0 -0.044361
10	brazil_varginha_T2M	2	0.039515	27	ethiopia_limu_WS2M	5 -0.058656
11	brazil_varginha_WS2M	5	-0.080176	28	ethiopia_sidamo_ALLSKY_SFC_SW_DWN	3 -0.068178
12	colombia_armenia_ALLSKY_SFC_SW_DWN	3	0.050835	29	ethiopia_sidamo_PRECTOTCORR	2 0.059995
13	colombia_armenia_PRECTOTCORR	1	0.052923	30	ethiopia_sidamo_T2M	6 0.067856
14	colombia_armenia_T2M	6	0.205598	31	ethiopia_sidamo_WS2M	6 -0.040277
15	colombia_armenia_WS2M	0	-0.055254	32	ethiopia_yirgacheffe_ALLSKY_SFC_SW_DWN	2 -0.101738
16	colombia_manizales_ALLSKY_SFC_SW_DWN	0	-0.028145	33	ethiopia_yirgacheffe_PRECTOTCORR	2 0.040343
				34	ethiopia_yirgacheffe_T2M	6 0.073674
				35	ethiopia_yirgacheffe_WS2M	1 -0.051427

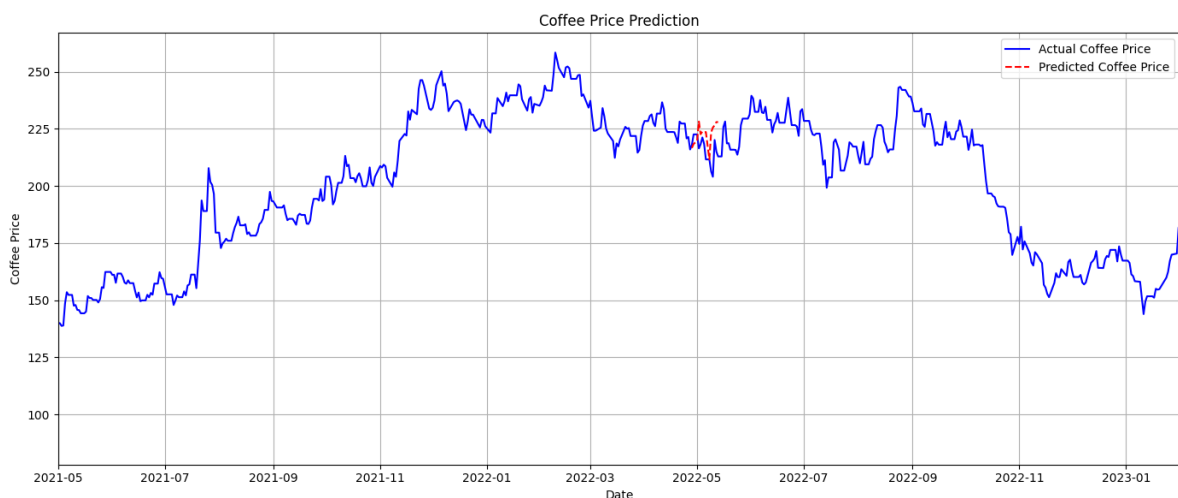
기후데이터에서 변수별로 lag를 현재시점~6개월전까지 적용하여 변수별 커피가격과 가장 높은 상관관계가 높은 lag값과 그때의 상관계수를 구해보았는데 상관계수의 절댓값이 많아도 0.2정도이다. 기존 모델에 기후데이터를 추가했을 때의 결과와 해당 상관관계의 결과를 확인해보았을 때 단순히 기후 데이터를 넣는다고 성능이 좋아지는 것은 어려울 수 있다.

기후데이터를 모델 학습에 사용하기 보다는 보조 지표로 만들어 모델 학습 마지막에 해당 지표를 통해 최종 예측을 하는 방식으로 하는 것이 더 좋을 것 같음

2. 기후 데이터를 통한 스코어를 만들어 보조지표로 사용

보조 지표를 만든다면 밑의 사진처럼 기준을 잡고 input date를 기준으로 최근에 harvest였을 때의 피처별 평균값을 이용해 스코어링하는 피처를 만들고 이를 최종 예측 결과에 반영하는 식으로 하는 방향이 있음

- Train 데이터의 마지막 date기준으로 2주후 후까지를 예측



- 모델 평가

MAE	RMSE
-----	------

3.74	4.31
------	------

기후 데이터는 가격에 직접적이지 않고 **비선형적, 간접적인 영향** 경로를 가질 가능성이 높아 **단순 피쳐로 사용할 경우 LSTM 기반 시계열 모델 성능을 향상시키기 어렵다.**

보조지표로서의 활용 가능성은 있으나, 현재로서는 성능 향상의 효과를 보장하지 못하였다. 기존 거시경제만 이용했을 때에 비하여

보조지표로서의 활용을 더 보강하거나 기후 -> 생산량 -> 가격의 간접 경로 모델링으로 접근하는 것이 더 효과적으로 보인다.