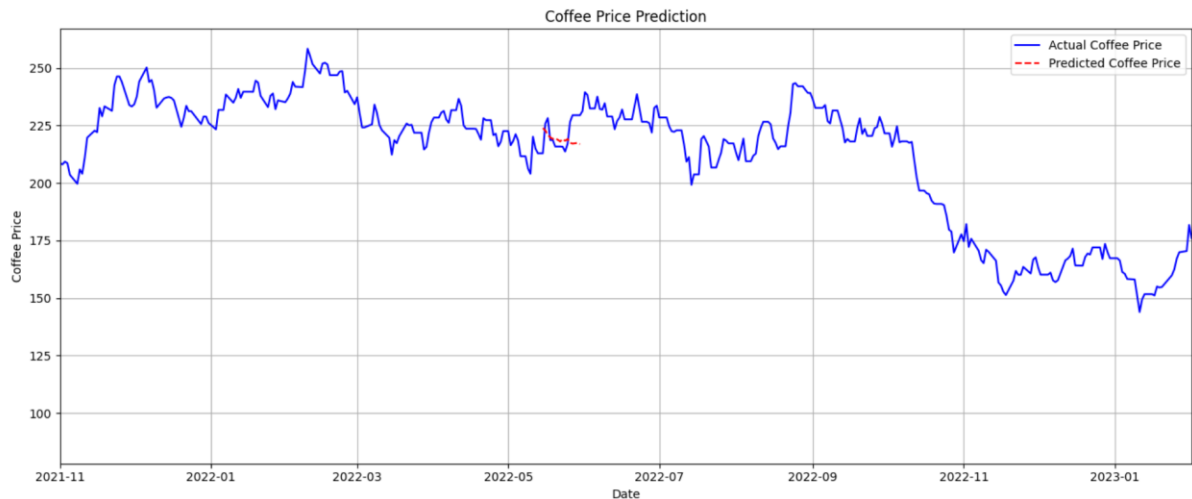


기후 데이터 사용 방법

1. 가장 최근 수확기의 기후데이터의 평균 값을 정적 피처로 사용

수확기 평균 기후는 그 해 생산된 커피의 품질을 대변한다. 즉, 시장에 판매되는 동안 계속 동일한 특성을 가지므로 가격에 지속적 영향을 미치는 고정적 특성으로 간주한다.

<예측/실제 비교 그래프>



<성능 평가>

MAE	RMSE
6.0631	7.4561

2. (1)의 모든 피처가 아닌 특정 피처만 사용

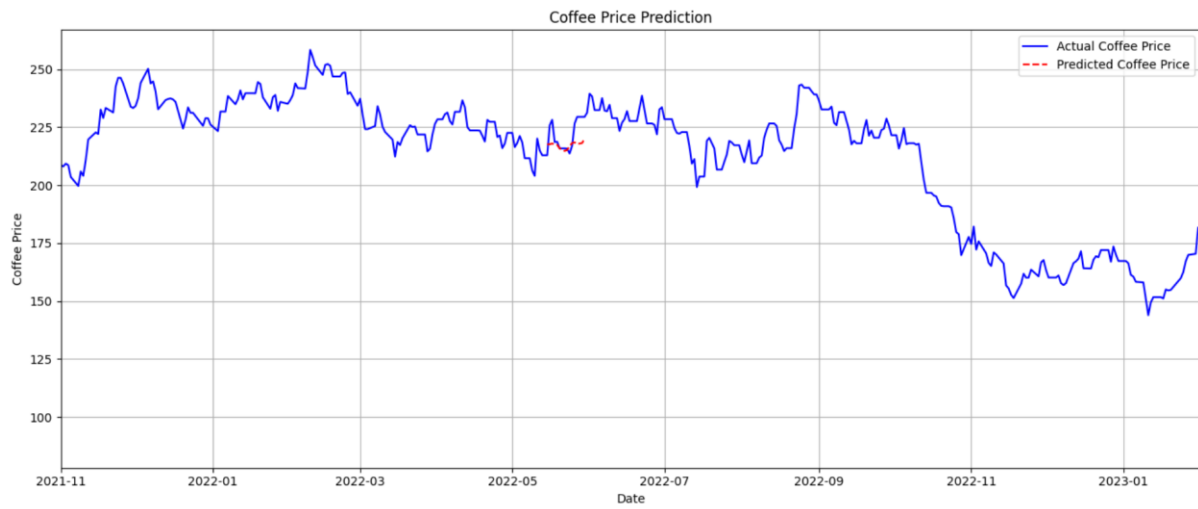
다수의 정적 피처가 들어가는 것 보다 LightGBM을 통해 feature importance만을 추출하여 중요도가 높은 정적피처만을 골라 학습에 사용하는 방법을 해보았다.

feature importance	
brazil_varginha_T2M_harvest_mean	417
colombia_manizales_T2M_harvest_mean	203
ethiopia_limu_T2M_harvest_mean	162
brazil_varginha_WS2M_harvest_mean	71
colombia_manizales_WS2M_harvest_mean	70
brazil_carmo_de_minas_PRECTOTCORR_harvest_mean	51
ethiopia_limu_PS_harvest_mean	45

LightGBM은 모델이 학습하는 동안 특정 feature가 얼마나 많이 결정에 사용되었는지를 측정하므로 중요도가 높을수록 해당 feature가 target feature의 예측에 기여한 정도가 크다는 것을 의미한다.

따라서 정적 피처 중 해당 피처들만을 사용해서 커피가격 예측을 실시하였다.

<예측/실제 비교 그래프>



여전히 급격하게 상승하는 부분에 대한 예측값은 실제값과의 차이가 크지만, 모든 정적 피처를 다 사용했을 때 보다 조금 더 변동성 있게 가격을 예측하고 있다.

<성능 평가>

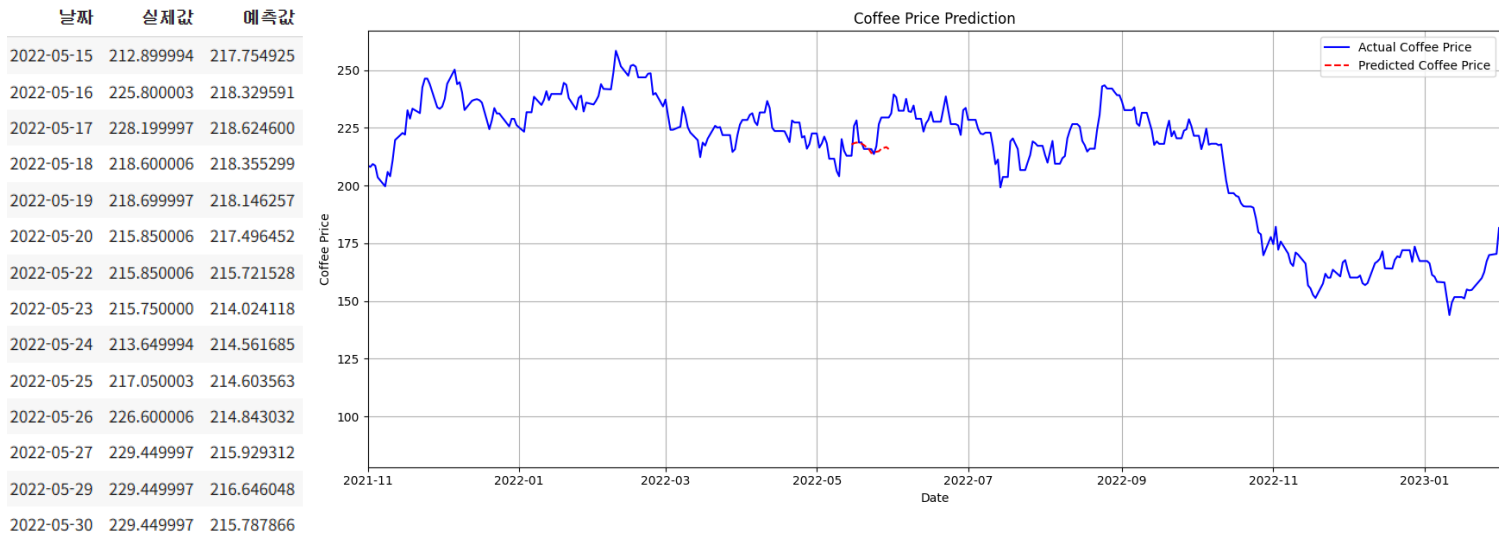
MAE	RMSE
5.1393	6.6995

Attention 사용

1. LSTM에 Attention 추가

각 timestamp의 출력이 attention의 대상이 된다. attention에서는 각 시점의 hidden state에 대해 score를 계산하여 softmax로 score를 확률처럼 변환하여 모든 시점에 weight를 부과한다. Weighted sum을 통해 시계열 요약 정보인 context vector를 계산한다.

<예측/실제 비교 그래프>



<성능 평가>

MAE	RMSE
5.8073	7.8175

전반적인 오차는 LSTM만 사용했을 때에 비해 높다. Softmax attention의 평준화 효과로 인해 예측이 전반적으로 평탄하게 유지되며 변동성이 줄어드는 문제가 발생한다. 지금 구조에서는 LSTM 단독 모델을 사용하는 것이 더 실용적인 선택이다. 따라서 일단 softmax를 사용하는 기존 attention이 아닌 변동성이 있는 시점에 집중할 수 있도록 하는 방식으로 개선한 후 결과를 확인해보았다.

2. softmax가 아닌 entmax 사용

위 처럼 일반 Attention은 softmax를 써서 모든 timestamp에 weight를 부과하기 때문에 평균화가 된다. Entmax는 softmax의 대안으로 출력 벡터 중 일부 값은 0이 되도록 만든 함수이다. 이를 통해 전체의 입력 중 모든 시점에 weight를 부여하지 않고 중요하게 생각하는 시점에만 weight를 부여해 핵심 구간에만 집중하게 만들 수 있다.

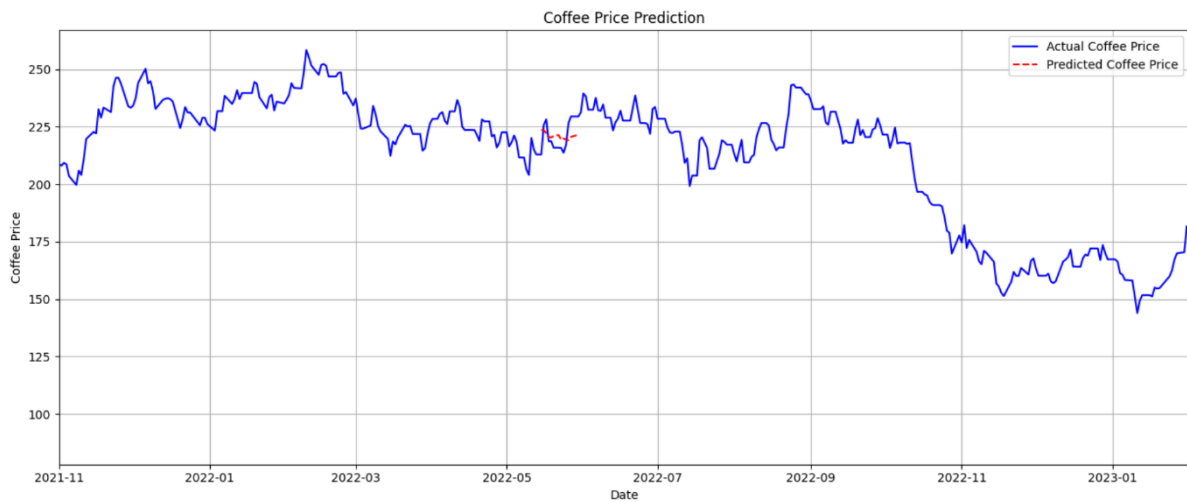
Softmax: 모든 입력에 양의 weight를 부여 → **dense (모두 반영됨)**

Entmax: 중요하지 않은 입력에는 **0을 출력함** → **sparse (선택적 반영)**

따라서 Entmax를 사용하면 평탄화된 예측을 줄이고 변동성있는 예측을 반영할 수 있게 된다.

<예측/실제 비교 그래프>

날짜	실제값	예측값
2022-05-15	212.899994	223.605186
2022-05-16	225.800003	223.433079
2022-05-17	228.199997	222.254007
2022-05-18	218.600006	220.568161
2022-05-19	218.699997	220.257883
2022-05-20	215.850006	220.969707
2022-05-22	215.850006	221.328873
2022-05-23	215.750000	219.320049
2022-05-24	213.649994	219.895094
2022-05-25	217.050003	219.321922
2022-05-26	226.600006	219.164470
2022-05-27	229.449997	220.543394
2022-05-29	229.449997	221.226532
2022-05-30	229.449997	220.548941



위에 softmax를 사용했을 때와 비교했을 때 급등하는 5/26, 5/30에 entmax를 사용했을 때가 더 강하게 반응한다. 좀 더 급등에 민감하게 반응하며 softmax에 비해서 방향성과 크기가 더 근접하다.

<성능 평가>

MAE	RMSE
5.6212	6.3046

3. 변동성 피처 추가

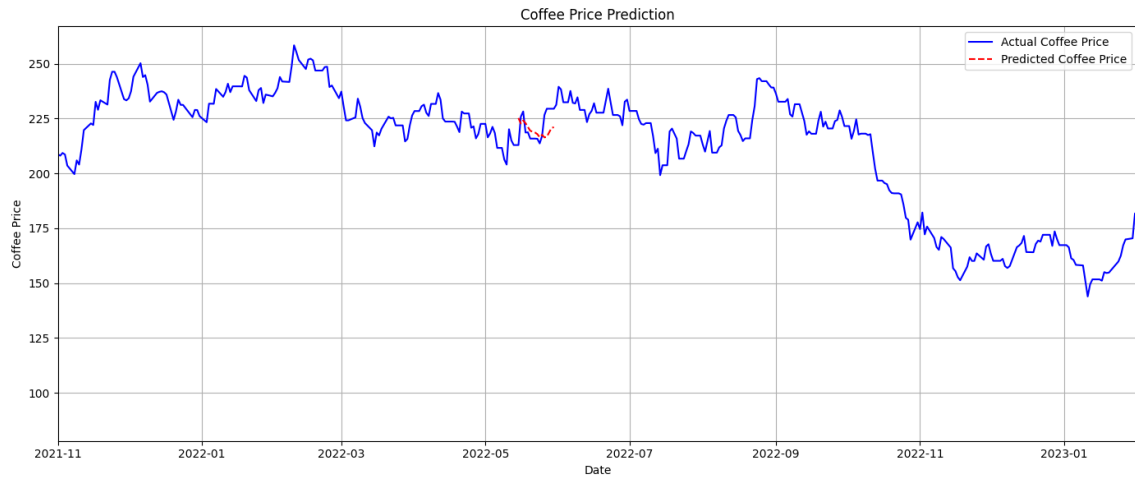
지금보다 더 커피 가격의 변동성을 반영하기 위해 변동성을 주기 위한 피처를 추가해주는 방식을 시도해보았다.

<추가 피처>

피처	생성 방법	정보
Abs_return	수익률의 절댓값	단기적 변화의 크기
volatility_n	n일간 수익률의 표준편차	변동성의 정도 측정
momentum_n	5일전 가격 대비 현재 가격	상승세, 하락세 감지
bollinger_width	볼린저 밴드 폭	시장이 과열인지 안정인지 측정
return_zscore	z-score 계산	이상치, 비정상적인 변동 감지

<예측/실제 비교 그래프>

날짜	실제값	예측값
2022-05-15	212.899994	225.201750
2022-05-16	225.800003	222.705843
2022-05-17	228.199997	224.113577
2022-05-18	218.600006	222.877582
2022-05-19	218.699997	221.116740
2022-05-20	215.850006	219.397214
2022-05-22	215.850006	218.619309
2022-05-23	215.750000	217.868534
2022-05-24	213.649994	216.441969
2022-05-25	217.050003	217.564867
2022-05-26	226.600006	216.389724
2022-05-27	229.449997	216.643847
2022-05-29	229.449997	220.697428
2022-05-30	229.449997	221.025391



<성능 평가>

MAE	RMSE
5.5794	6.8158

변동성 피처를 추가하였을 때 MAE와 RMSE 모두 개선되었다.

여전히 성능 평가 지표로 보았을 때는 MultiStepLSTM을 사용하였을 때보다 성능이 좋지 못하다. 하지만 예측 범위가 217~220 수준으로 평탄하였던 것에 비해 216~225로 폭이 넓어졌다. 즉 실제 값 방향성 반영이 더 잘 되고 있다. 하지만 여전히 5월 26, 27일에 급등하는 부분에 덜 민감하게 따라간다. 모델이 스무딩 된 예측을 하는 경향이 여전히 있으므로 급등락 반영을 위한 손실함수의 수정을 통해 실제 추세를 더 잘 반영하도록 하는 수정이 필요하다.

<현재 사용 데이터>

데이터	처리	최종 생성 데이터
거시경제데이터	스플라인 보간법을 사용해 연간 -> 일간	커피가격의 변동성 반영을 위한 보조피쳐 추가
기후데이터	가장 최근 수확기의 피쳐별 평균값	LightGBM을 통한 feature selection 수행

<현재 모델 설계 요약도>

구성 요소	방식	기능 및 목적
시계열 인코더	LSTM	순차 정보, 추세 학습
Attention 메커니즘	Entmax	주요 시점 집중
Context 조합 방식	Context + last hidden	유연한 정보 선택
예측 헤드	2-layer FC + ReLU	다양한 패턴 예측
Optimizer	Adam	안정적 학습

<모델 구조 흐름 요약>

