

데이터를 통해 사용자와 세상을 연결하는 인재, **나리나**입니다.



나리나 Na Rina

Phone +82 10-6381-3274

mail skery@naver.com

Github <https://github.com/Na-Rina>

Education

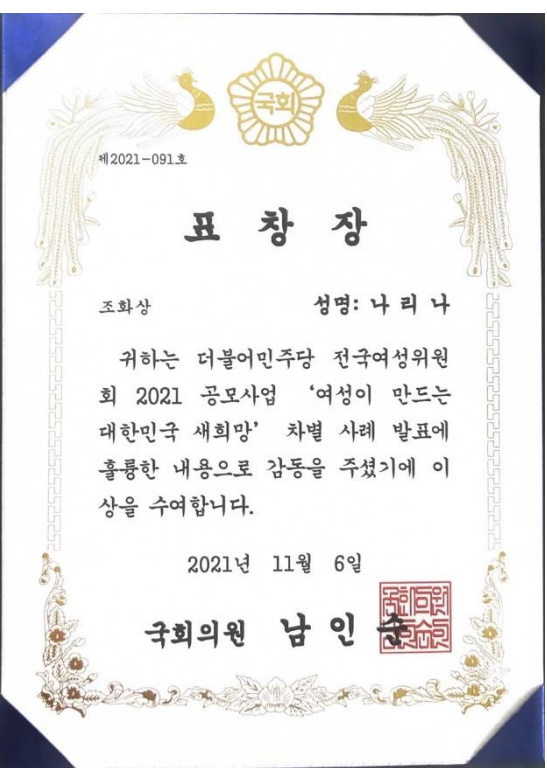
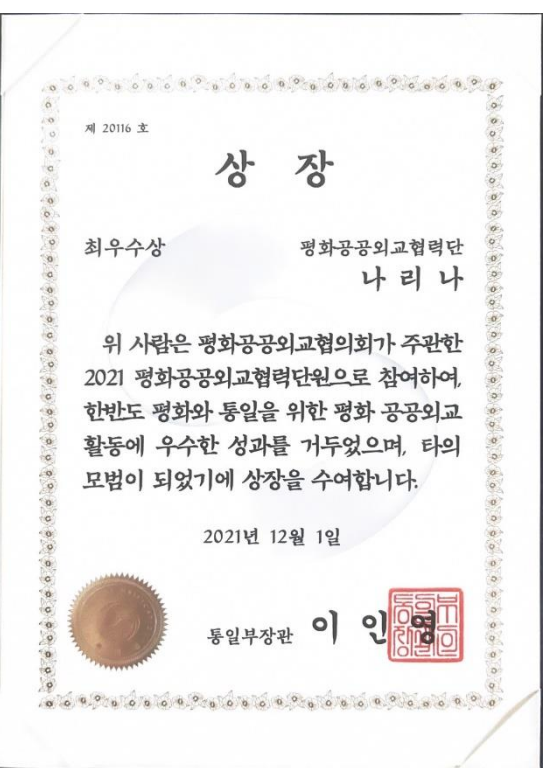
- 2020 오금고등학교 졸업
- 2025 한국외국어대학교 재학 중
 - 프랑스학과, AI & Software 융합 전공
- 2025 - 산업은행 디지털전략부 AI개발팀 AI인턴

Activities

- 2022. 01 ~ 2022. 12 한국외국어대학교 외대학보
- 2022. 03 ~ 2022. 12 ICT 봉사단 마다가스카르 대학교
- 2022. 05 ~ 2024. 06 송파부모연대 장애인활동지원사
- 2023. 09 ~ 2024. 05 한국심리상담복지학회 논문 작성

Certificate & Award

- 2020 컴퓨터활용능력 2급
- 2021 Flex (프랑스어 읽기, 쓰기)
- 2021 통일부 장관상 최우수상, 송파구 국회의원 남인순 우수상
- 2024 TOEIC 785
- 2024 Opic IH (Intermediate High)
- 2024 DAsP (데이터아키텍처전문가)
- 2024 ADsP (데이터분석준전문가)
- 2024 SQLD (SQL 개발자)
- 2024 Google Ads Search Certification
- 2024 Google 애널리틱스 individual Qualification
- 2025 한경협 ESG 전문가 자격증
- 2025 빅데이터분석기사



강화학습을 이용한 장바구니 추천 시스템 앱 개발 깃헙

프로젝트 개요 및 목표

신세계 API와 강화학습 모델을 결합해 사용자 맞춤형 장바구니 추천 시스템을 개발하고, AWS 기반의 안정적인 서버 환경에서 성능 극대화를 목표로 진행한 프로젝트입니다.



프로젝트 기간

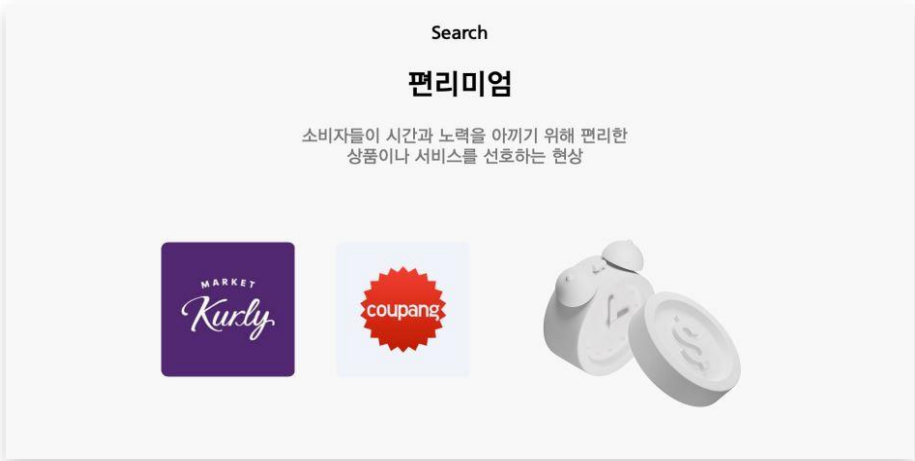
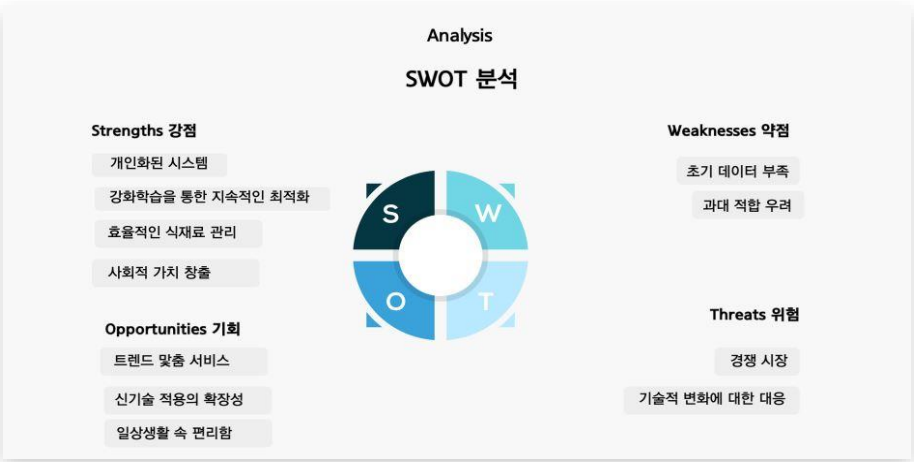
- 2024. 01 ~ 2024. 06 (5개월)

프로젝트 팀 구성

- 총 4명 : Frontend, Backend, DB, 머신 러닝 역할 분담

프로젝트 목표

- 신세계 API를 활용해 사용자 맞춤형 장바구니 추천 시스템 개발
- 강화학습 모델을 적용하여 추천의 정확도 향상
- AWS 클라우드 인프라로 안정적인 운영 환경 확보



강화학습을 이용한 장바구니 추천 시스템 앱 개발

시스템 구성 및 프로젝트 성과

프로젝트 기여도

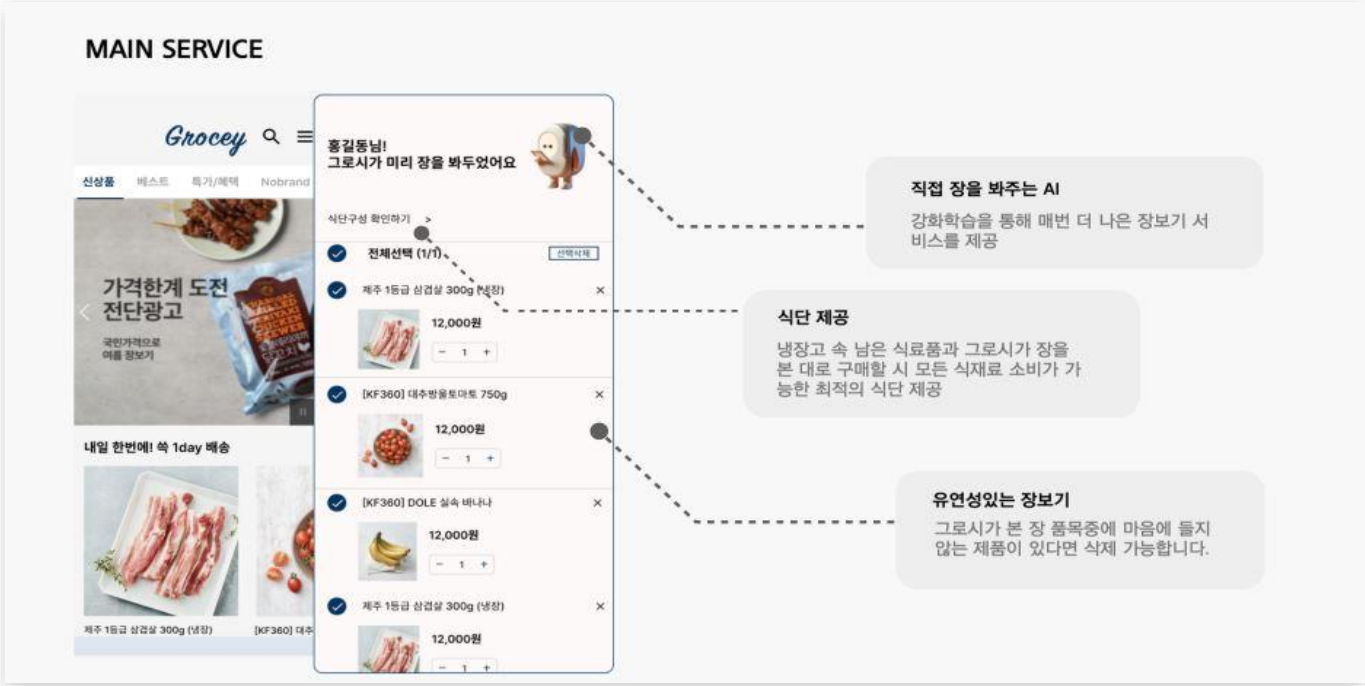
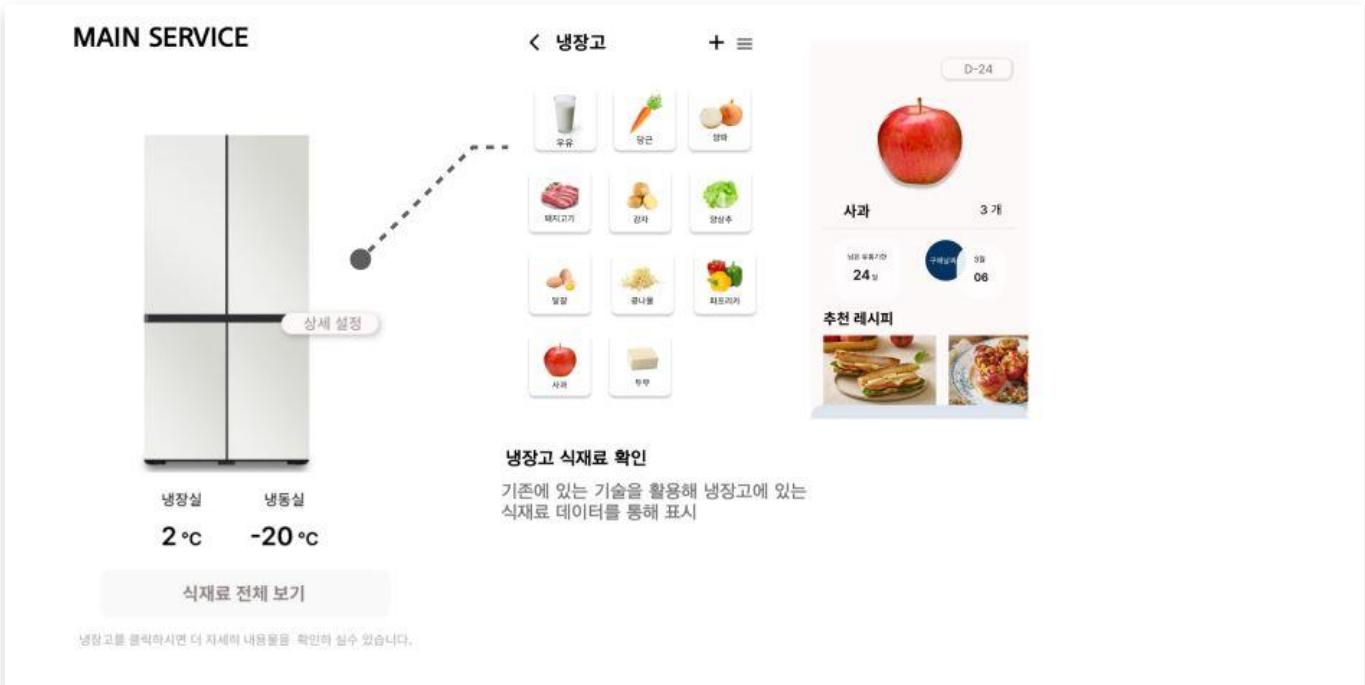


- 데이터 처리 : 더미 데이터 생성 및 전처리
- 모델 구현 : 강화학습 AI 모듈 개발
- AWS 활용 : EC2, RDS, S3 서버 구성 및 데이터 연계
- 문서화 및 발표 : 프로젝트 문서 작성 및 발표 자료 제작

* 기술 스택 : Python, AWS, Docker, Node js ...

프로젝트 성과

- 강화학습을 통해 추천 정확도 10% 개선
- AWS를 이용한 안정적인 클라우드 기반 배포
- 사용자 맞춤형 추천 시스템 초기 버전 구현 완료



WFSU 앱 개발 (공공 와이파이 보안 강화 및 이상 탐지 서비스) 깃합 | AI 보고서

프로젝트 개요 및 목표

서울특별시 공공 와이파이 데이터를 활용하여
AI 기반으로 비정상 AP를 탐지하고, QGIS를 통해 위험도를 시각화하여
디지털 약자의 안전한 와이파이 이용을 지원하는 서비스를 개발한 프로젝트입니다.

프로젝트 기간

- 2024. 03 ~ 2024. 08 (5개월)

프로젝트 팀 구성

- 총 4명 : Frontend, Backend, Data AI, 클라우드 역할 분담

프로젝트 목적

- 디지털 약자를 위한 공공 와이파이 보안 강화 및 이상 탐지
- 이상 AP(Access Point) 탐지 및 서비스 개선을 위한 지도 기반 시각화



주요 작업 목표

서울특별시 공공 와이파이 데이터를 분석 및 활용

- 이상 사용량을 탐지하여 위험 AP를 식별
- AI 기반 Isolation Forest 알고리즘으로 이상치(비정상 AP) 탐지

결과를 QGIS 기반으로 시각화

- 자치구별 공공 와이파이 AP 분포 정보 제공
- 이상치 AP를 위험 수준별 색상 구분으로 표시하여 직관적인 위험도 전달

WFSU 앱 개발 (공공 와이파이 보안 강화 및 이상 탐지 서비스)

프로젝트 결과 & 기술 스택

프로젝트 기여도

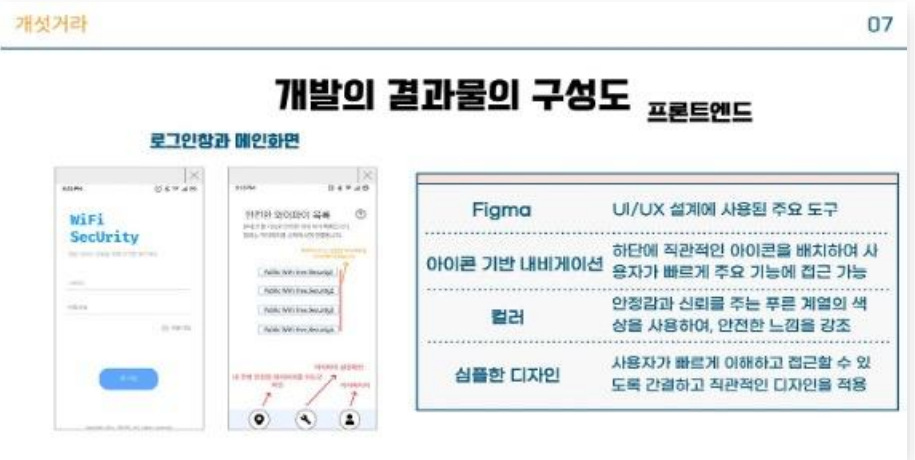


- 데이터 처리 : 서울 공공 와이파이 AP 데이터 수집 및 전처리
- AI 모델링 : 이상치 탐지 모듈 (Isolation Forest) 작성
- 지도 시각화 : QGIS 활용, 공공 와이파이 분포 및 위험도 통합 시각화
- 문서화 및 발표 : 최종 보고서 작성 및 결과 발표 진행

* 기술 스택 : Python, AI, GCP, QGIS

프로젝트 성과 : 사용자 보안을 강화한 서비스 설계 및 구현

- 와이파이 10,191건 중 Isolation Forest로 이상 AP 510건 탐지
- QGIS 지도 기반 시각화 : 자치구별 와이파이 분포를 맵핑 및 위험 AP를 색상으로 구분
- GCP를 활용한 안정적 데이터 처리 및 분석 환경 구축
- 디지털 약자를 위한 안전 와이파이 사용 인식 개선 및 경고 시스템 제공



농산물 데이터 예측(LSTM) 깃헙

프로젝트 개요 및 목표

본 프로젝트는 캐글 "제주 특산물 가격 예측 AI 경진대회" 데이터를 활용하여 LSTM 기반 예측 모델을 개선하고, 평균 제곱 오차(MSE)를 약 74% 감소시키며 농산물 가격 예측의 정확성을 크게 향상시켰습니다.

프로젝트 기간

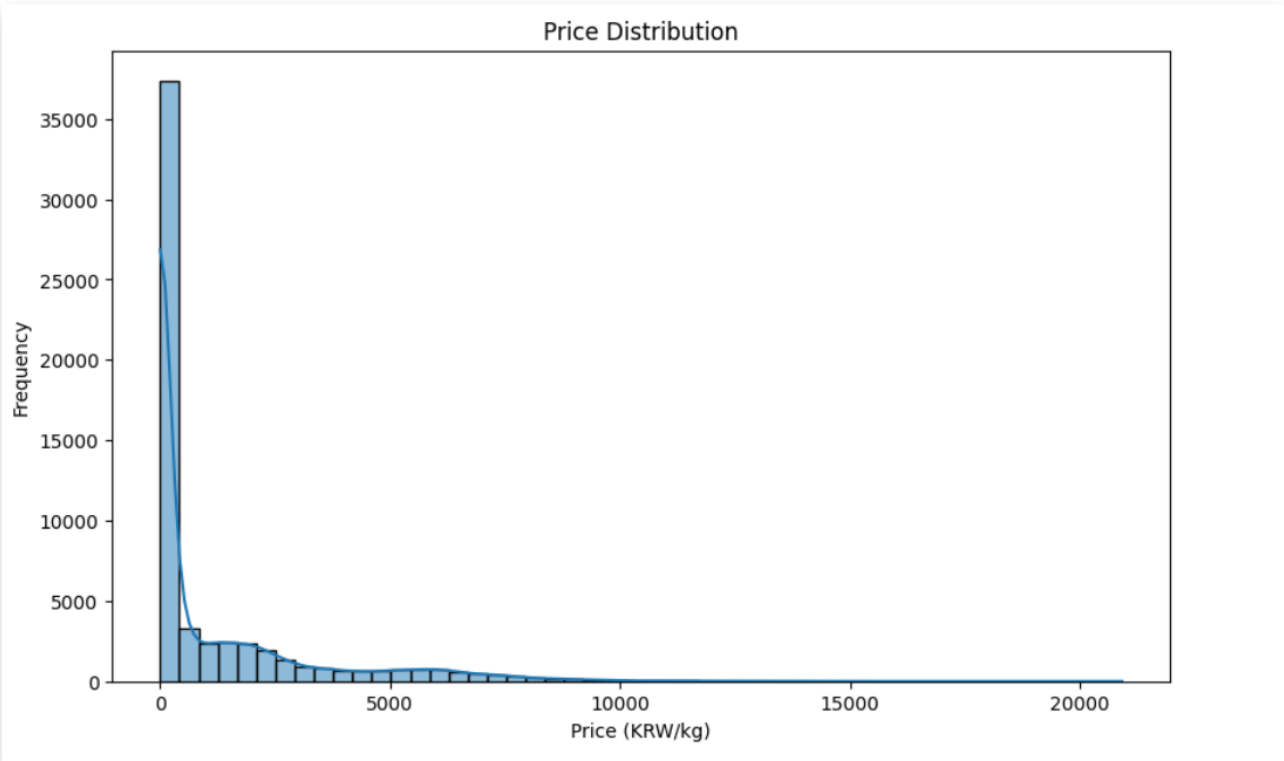
- 2023. 12 ~ 2024. 07 (8개월)

프로젝트 목표

- 캐글 "제주 특산물 가격 예측 AI 경진대회" 데이터를 활용하여 정확도가 높은 농산물 가격 예측 모델을 개선

프로젝트 내용

- GPU 기반 학습: 디바이스 설정을 통해 학습 속도와 효율성을 향상.
- 손실 함수: 평균 제곱 오차(MSE)를 활용하여 예측 오차를 최소화.
- 옵티마이저: Adam 옵티마이저로 학습 안정성과 최적화 효과를 극대화.



농산물 데이터 예측(LSTM)



코드 설명 영상

프로젝트 성과 및 세부 사항

프로젝트 결과

- 모델 성능 향상

모델 구조 변경 및 하이퍼파라미터 튜닝으로 예측 정확도 대폭 향상

평균 제곱 오차(MSE)가 크게 감소 (약 74% 감소).

* 기술 스택 : Python, RNN, LSTM, AutoML

	기존 모델	개선된 모델
모델 구조	입력크기 50, 레이어 수 1	RNN · LSTM 기반 다층 구조
MSE	2,177,767.0118	565,082.1202



본 프로젝트를 통해 농산물 가격 예측 모델의 신뢰성과 활용 가능성을 크게 증대

	ID	timestamp	item	corporation	location
0	TG_A_J_20230304	2023-03-04	TG	A	J
1	TG_A_J_20230305	2023-03-05	TG	A	J
2	TG_A_J_20230306	2023-03-06	TG	A	J
3	TG_A_J_20230307	2023-03-07	TG	A	J
4	TG_A_J_20230308	2023-03-08	TG	A	J
...
1087	RD_F_J_20230327	2023-03-27	RD	F	J
1088	RD_F_J_20230328	2023-03-28	RD	F	J
1089	RD_F_J_20230329	2023-03-29	RD	F	J
1090	RD_F_J_20230330	2023-03-30	RD	F	J
1091	RD_F_J_20230331	2023-03-31	RD	F	J

1092 rows × 5 columns

분포 그래프 분석

그래프 설명

위의 그래프는 'Train' 데이터셋의 'price(원/kg)' 열에 대한 가격 분포를 보여줌. 가격은 한국 원(KRW) 단위로 측정되었으며, x축은 가격, y축은 빈도수를 나타냄.

분석 결과

1. 가격 분포의 집중:

- 그래프에서 볼 수 있듯이, 가격이 0인 데이터 포인트가 매우 많이 분포하고 있음. 이는 공급량이 0이거나 가격 정보가 누락된 경우일 수 있음.
- 대다수의 데이터 포인트는 저가 구간에 몰려 있으며, 특히 0에 매우 높은 빈도수가 집중되어 있음.

2. 긴 꼬리 현상:

- 그래프는 오른쪽으로 긴 꼬리를 가지고 있으며, 이는 고가 데이터 포인트가 존재하지만 빈도가 매우 낮음을 나타냄.
- 대부분의 가격이 0에서 5,000 KRW/kg 사이에 분포되어 있으며, 그 이후의 가격은 상대적으로 적음.

3. 가격 변동성:

- 가격 분포의 표준편차가 크다는 것을 알 수 있음. 이는 가격 데이터의 변동성이 크다는 것을 의미함.
- 저가부터 고가까지 넓은 범위의 가격이 존재함.

결론

데이터 전처리 필요성:

- 가격이 0인 데이터 포인트에 대한 추가적인 분석이 필요함. 이 데이터들이 실제로 의미 있는 값인지, 아니면 오류나 누락된 데이터인지를 고려해야 함.
- 가격 변동성이 큰 데이터의 특성을 고려하여 모델링 시 적절한 스케일링이나 변환을 적용하는 것이 필요할 수 있음.

모델링 전략:

- 저가 구간에 데이터 포인트가 집중되어 있으므로, 모델링 시 이 부분에 대한 정확한 예측이 중요함.
- 고가 구간의 데이터 포인트는 적지만 중요한 정보를 포함할 수 있으므로, 이들을 적절히 다룰 필요가 있음.

하이퍼파라미터 튜닝

```
In [37]: input_size = X.shape[1] # 특성 수
hidden_size = 100 # 은닉 상태 크기 증가
num_layers = 2 # LSTM 레이어 수 증가
output_size = 1

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model = ImprovedModel(input_size, hidden_size, num_layers, output_size).to(device)
criterion = nn.MSELoss()
optimizer = torch.optim.Adam(model.parameters(), lr=0.001)

# 학습 루프
epochs = 200 # 에포크 수 증가
for epoch in range(epochs):
    model.train()
    for inputs, targets in train_loader:
        inputs, targets = inputs.to(device), targets.to(device)
        outputs = model(inputs.unsqueeze(1))
        loss = criterion(outputs, targets.unsqueeze(1))
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()

    print(f"개선된 모델 학습 완료")
```

개선된 모델 학습 완료

```
In [39]: # 개선된 모델 평가
improved_model_loss = evaluate_model(model, val_loader)
print(f"개선된 모델 MSE: {improved_model_loss:.4f}")
```

개선된 모델 MSE: 565082.1202

5. 성능 비교 및 결론

기존 모델

- 은닉 상태 크기: 50
- LSTM 레이어 수: 1
- 학습률: 0.001
- 에포크 수: 100
- 평균 제곱 오차 (MSE): 2177767.0118

개선된 모델

- 은닉 상태 크기: 100 (은닉 상태 크기를 증가시켜 모델의 용량을 확장)
- LSTM 레이어 수: 2 (더 많은 레이어를 추가하여 모델의 깊이를 증가)
- 학습률: 0.001 (기존과 동일)
- 에포크 수: 200 (더 많은 에포크 수를 사용하여 모델을 더 오래 학습)
- 평균 제곱 오차 (MSE): 565082.1202

결론

- 모델 구조 변경: 더 깊은 LSTM 레이어와 더 큰 은닉 상태 크기를 사용한 개선된 모델은 기존 모델에 비해 더 나은 성능을 보였습니다.
- 하이퍼파라미터 튜닝: 에포크 수를 늘려 더 오래 학습시킨 결과, 개선된 모델의 예측 성능이 크게 향상됨
- 평가 결과: 개선된 모델의 MSE가 기존 모델의 MSE보다 크게 낮아, 데이터의 패턴을 더 잘 학습했음을 알 수 있음

이와 같은 개선을 통해 LSTM 모델의 예측 성능을 효과적으로 향상시킬 수 있음을 확인할 수 있었음