



암호화폐 예측 모델링

박종민, 이시준

목차

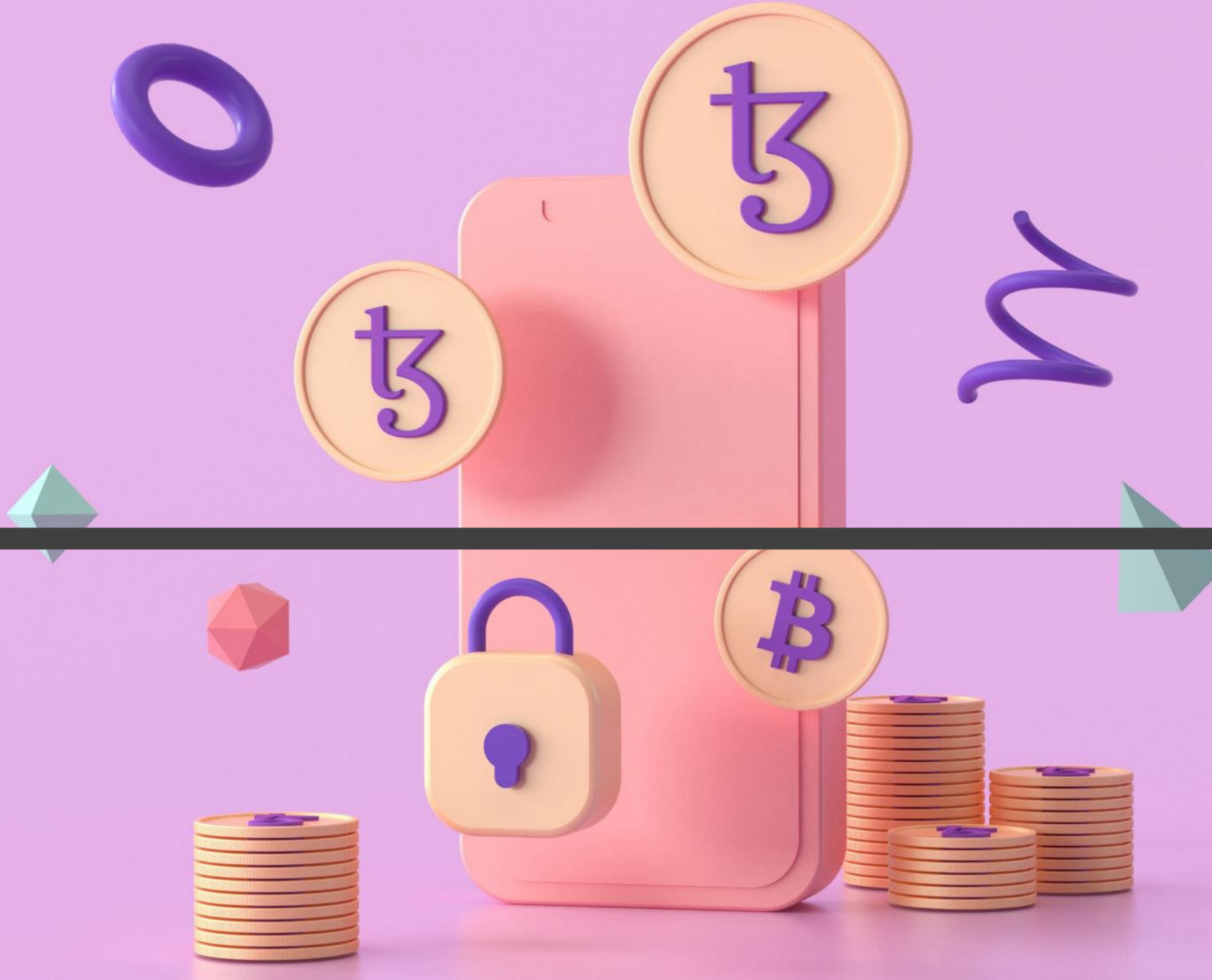
1 주제 선정

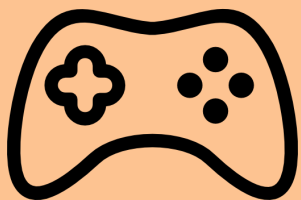
2 모델링

3 결과

Part 1

주제 선정

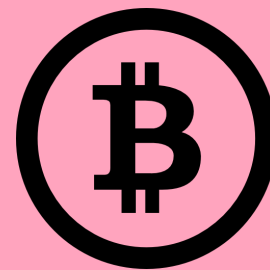




LoL 실시간 승률 예측



기사 정치 편향성 판단 모델링



비트코인 예측 모델링

인공신경망 모델을 이용한 뉴스 기사 분석에 의한 주가지수 예측 방법 및 장치

Abstract

본 발명은 인공신경망 모델을 이용한 뉴스 기사 분석에 의한 주가지수 예측 방법 및 장치에 관한 것으로서, 더욱 상세하게는 다수의 뉴스 기사를 기계학습(machine learning)된 인공신경망 모델에 입력하여, 이 모델로부터 매일의 주가지수를 예측하는 방법 및 그 방법을 수행하는 장치에 관한 것이다.

본 발명에 의하면, 뉴스 기사로부터, Open IE에서와 같이 주어(S), 동사(V), 목적어(O)와 유사한 구조를 갖는 relation tuple(O1; P; O2)를 추출하지 않고 이보다 낮은 단위인 단어 레벨 정보를 추출함으로써, 정보 손실을 줄인 입력을 인공 신경망으로 제공한다. 뉴스 기사 타이틀은 비교적 짧으며, 사용된 단어 수 편차가 적은 데이터이므로, 이와 같이 relation tuple의 형태를 적용하지 않고 단어 레벨 정보를 입력으로 사용함으로써 정보 손실을 훨씬 줄이고 유용한 결과를 내게 한다. 또한 인공신경망으로서 RNN(recurrent neural network) 또는 RNN의 일종인 LSTM(long short-term memory) 모델을 사용함으로써 뉴스 기사의 단어와 같이 순차적으로 등장하는 데이터 처리를 더욱 정확하게 하여, 주가지수에 대하여 더욱 신뢰성있는 결과를 도출할 수 있도록 한다.

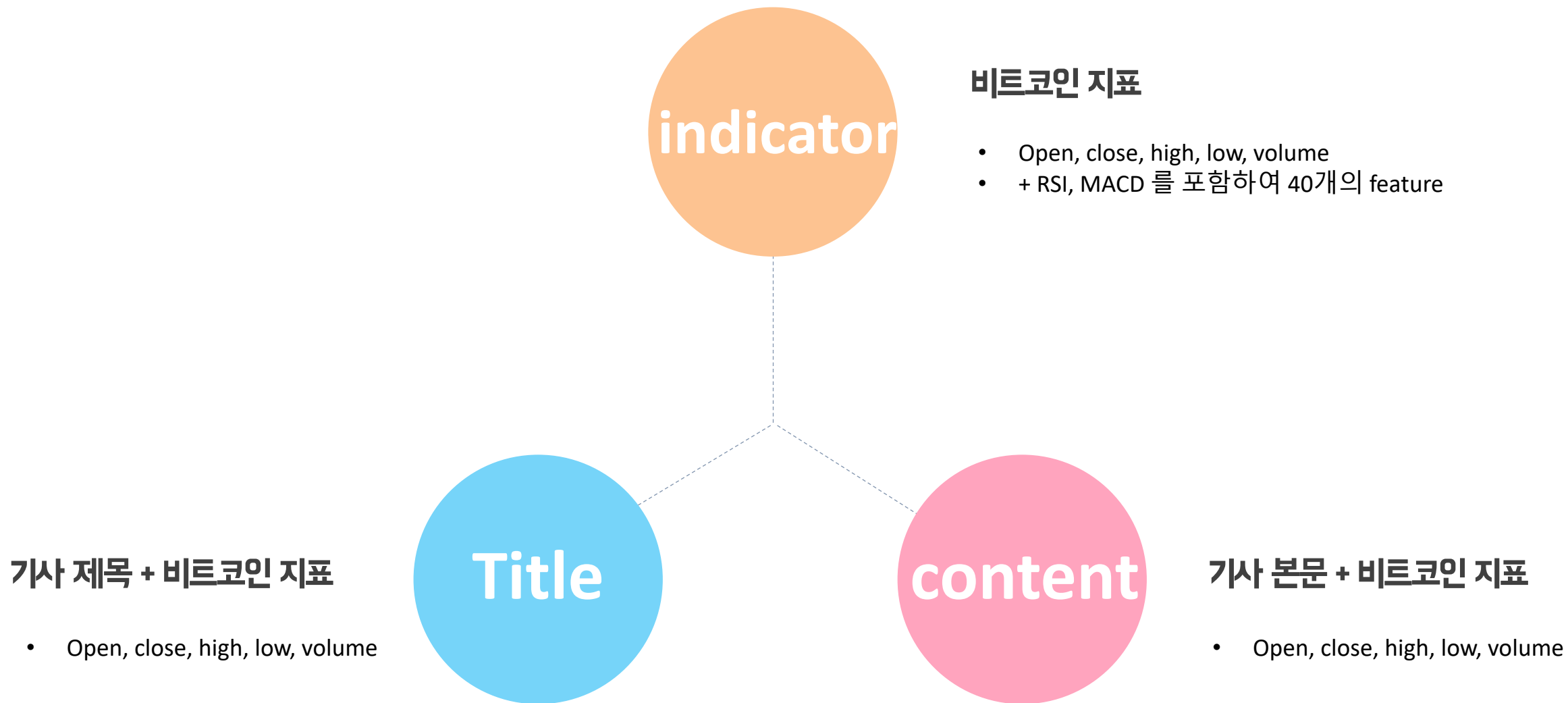


#BITCOIN

Part 2

모델링





기사	비트코인
<p>네이버 검색어 “비트코인”을 중심으로 2017.9.30 ~ 2023.5.23일까지의</p> <p>총 140,060개의 기사 (크롤링)</p>	<p>2017.9.30 ~ 2023.5.23일까지의 5분단위 비트코인 지표 데이터 (출처 : Kaggle)</p> <p>약 580,000개의 row</p>

“비트코인” 검색어를 중심으로

기사 크롤링

기사 임베딩

Transformer 기반 시계열 예측 모델

모델 훈련

금융단어 특화 BERT 사전학습모델
(kakao bank 개발)

주어진 feature 기반으로 증감
+ 금 거래량, 금리, snp 지수

데이터 feature 증감
(5 → 40)

데이터 전처리

결측치 제거, 스케일링(MinMax)

RNN 기반 시계열 예측 모델

모델 훈련

뉴스기사 모델링의 한계점

1

뉴스 기사 하나가 실제 비트코인 가격에
영향을 주는 정도가 확실하지 않다.

2

기사를 전처리 했을때,
기사가 나오지 않는 시간대가 많다.

3

기사 임베딩 모델 성능의 한계
>>kf-deberta-base

4

뉴스 기사가 영향을 미치는
시간대를 통일 할 수 없다.
(기사 발행 후 5분 후, 10분 후,...)

비트코인 지표

- Bi GRU

indicator

indicator

비트코인 지표

- Bi LSTM

Final
model

비트코인 지표

- Bi GRU



indicator

Bi GRU

(Bidirectional Gated Recurrent Unit)

장점

- **문맥 이해 향상:** 양방향으로 데이터를 처리하여 앞뒤 맥락을 모두 반영할 수 있어 성능이 좋다.
- **효율성:** GRU의 단순한 구조 덕분에 계산 비용이 상대적으로 낮다.

Bi LSTM

(Bidirectional Long Short-Term Memory)

장점

문맥 이해 향상: 양방향으로 데이터를 처리하여 앞뒤 맥락을 모두 반영할 수 있어, 성능이 좋다.

장기 의존성 처리: LSTM의 구조 덕분에 긴 시퀀스에서도 중요한 정보를 잃지 않는다.



비트코인 지표

- Bi LSTM

가중평균

Validation loss를 기준으로
두 개의 모델을 가중평균으로 최종 예측 가격을 도출한다.



Final
model

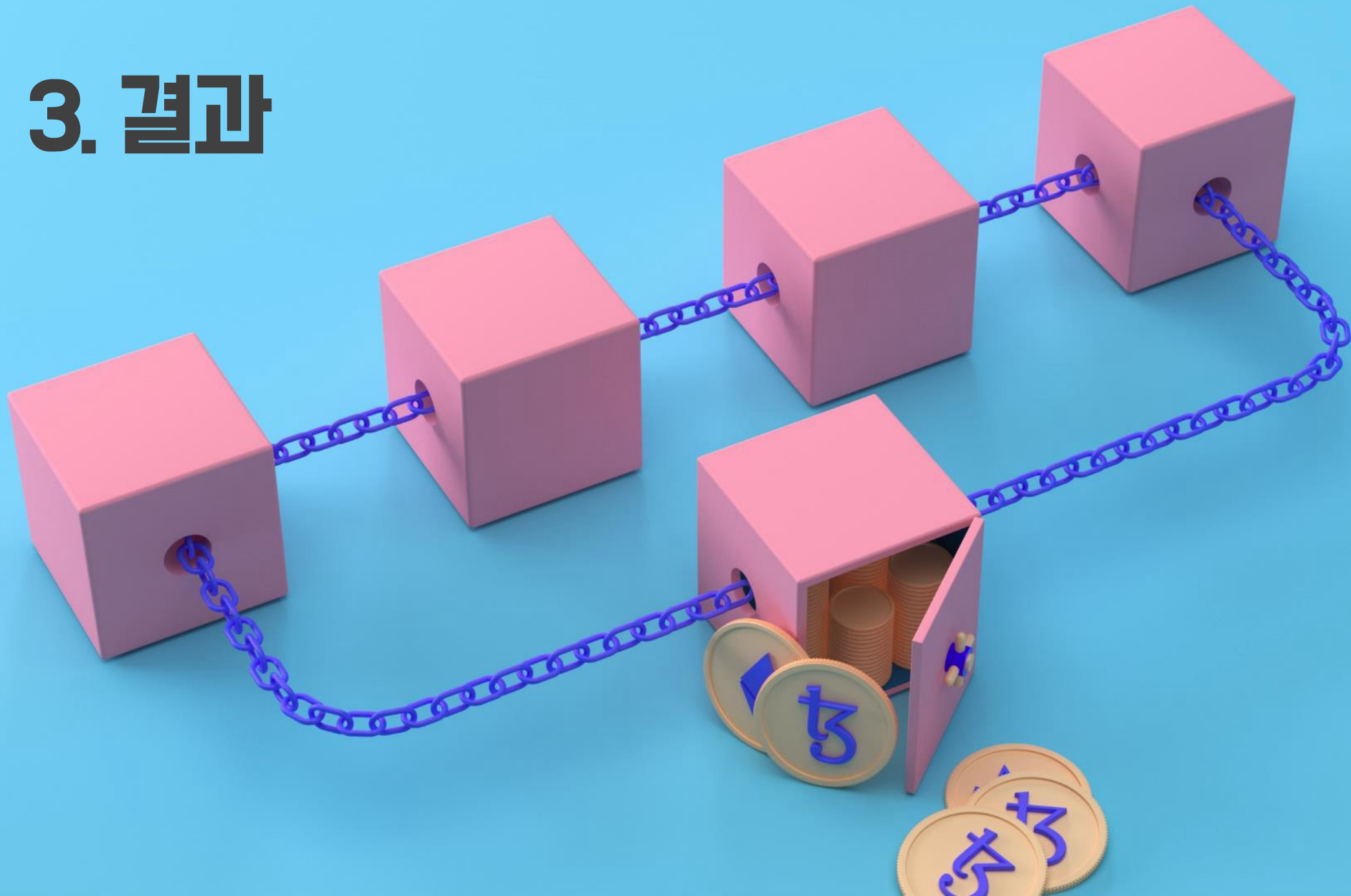
1 Bi-GRU

2 Bi-LSTM

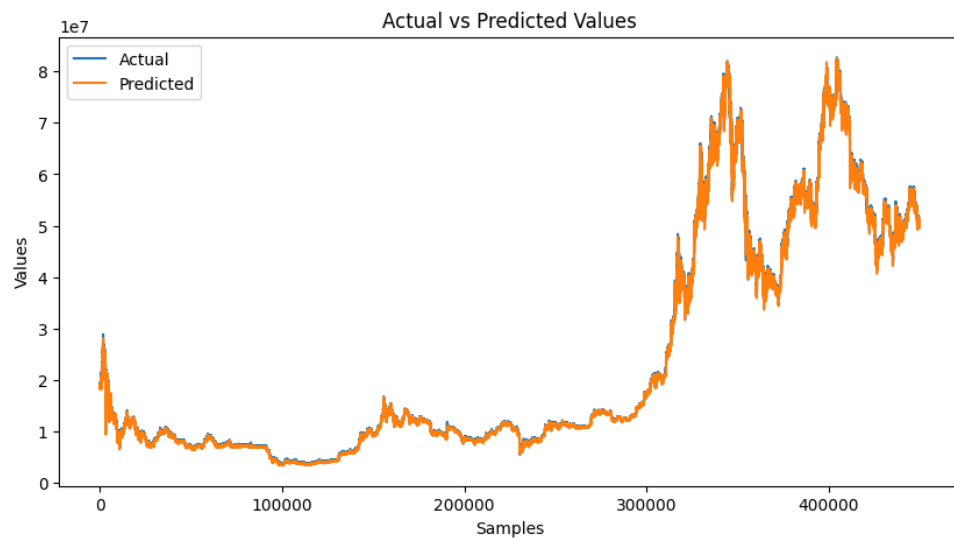
3 Final - Model

하이퍼파라미터를 공개하고 싶었으나,
수익률이 생각보다 높아 투자만 받겠습니다.

3. 결과

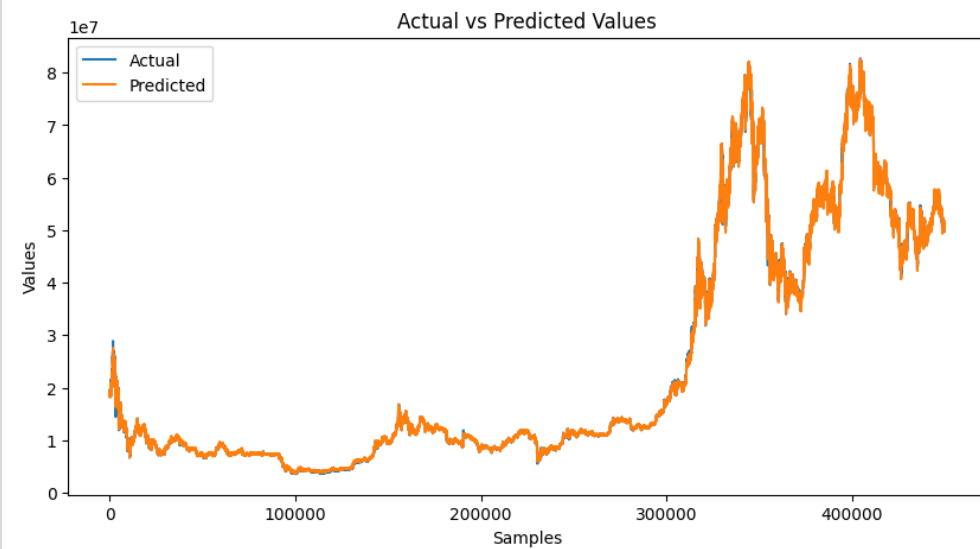


1 Bi-GRU



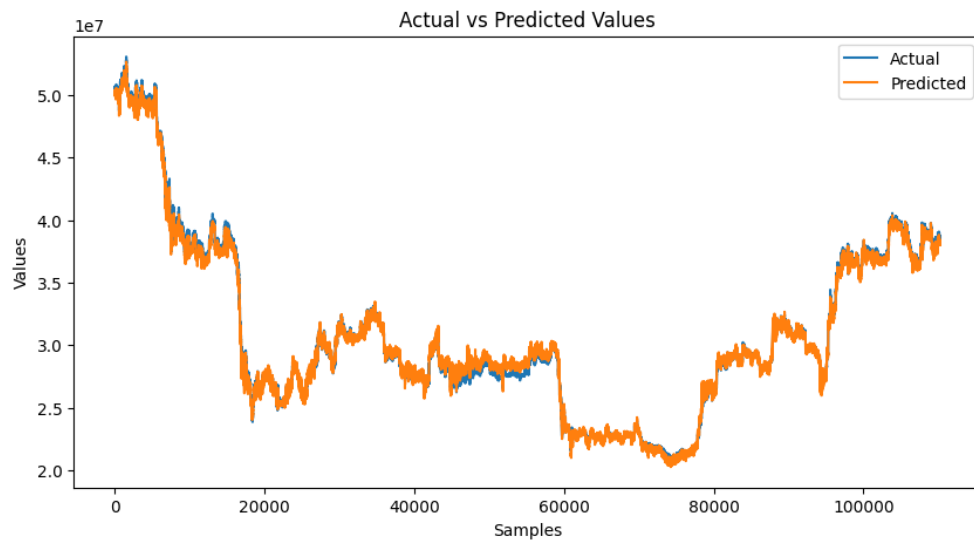
Loss : 0.000175

2 Bi-LSTM



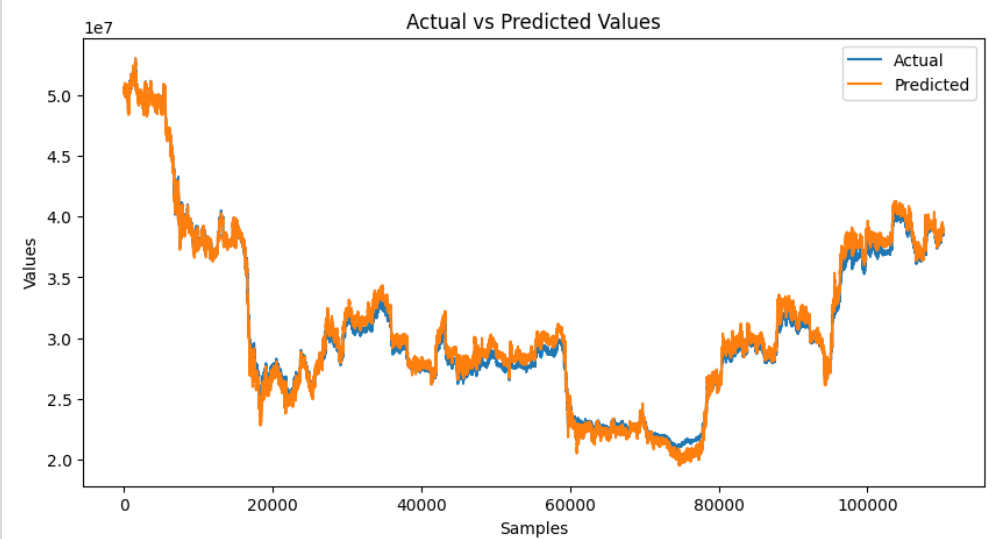
Loss : 0.000209

1 Bi-GRU



Loss : 0.000021

2 Bi-LSTM



Loss : 0.000061

Back-testing 모의투자

Bi-GRU	Bi-LSTM	Final model
기간 : 2023년 Seed : 1,000,000,000원(10억)		
Final cash 1,875,465,368원 (약 19억)	Final cash 1,191,535,690원 (약 12억)	Final cash 2,054,507,142원 (약 20억)
수익률 : 87.5% (+8억 7천)	수익률 : 11.9% (+2억)	수익률 : 105% (+10억 5천)



Part 4 Challenges

Modeling error 1

```
Epoch [1/20], Train Loss: 1042862323323015.1250, Val Loss: 1007847434035979.2500
```

```
Epoch [2/20], Train Loss: 1042851179774231.6250, Val Loss: 1007833485746651.1250
```

```
Epoch [3/20], Train Loss: 1042840484136198.1250, Val Loss: 1007819448880629.2500
```

```
Epoch [4/20], Train Loss: 1042829784109116.6250, Val Loss: 1007805413649510.1250
```

```
Epoch [5/20], Train Loss: 1042819093825501.0000, Val Loss: 1007791467092400.3750
```

```
Epoch [1/20], Train Loss: 1033662951944522.8750, Val Loss: 990281809360739.2500
```

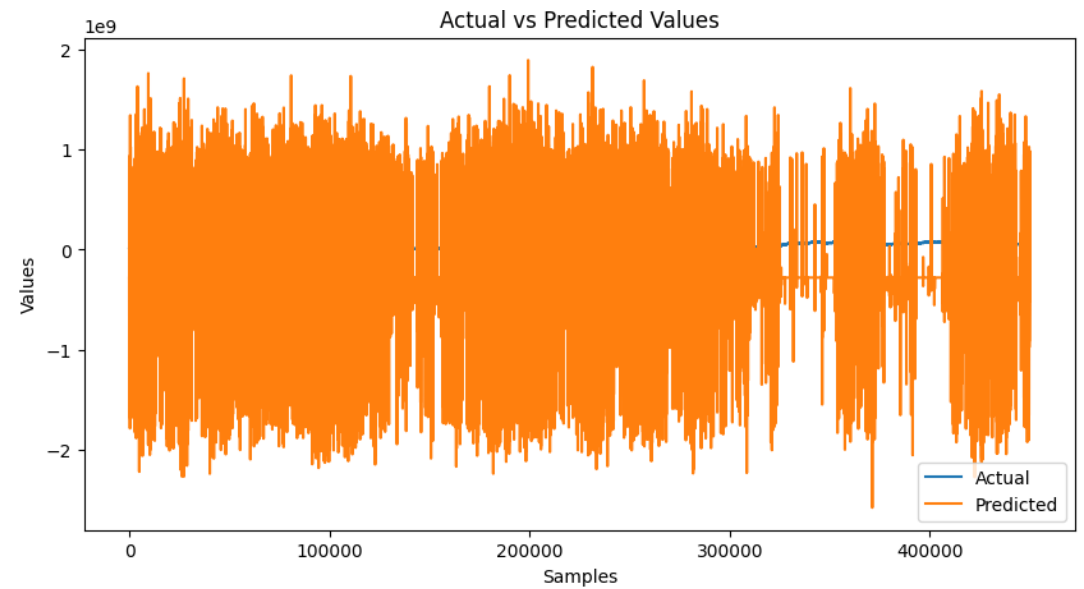
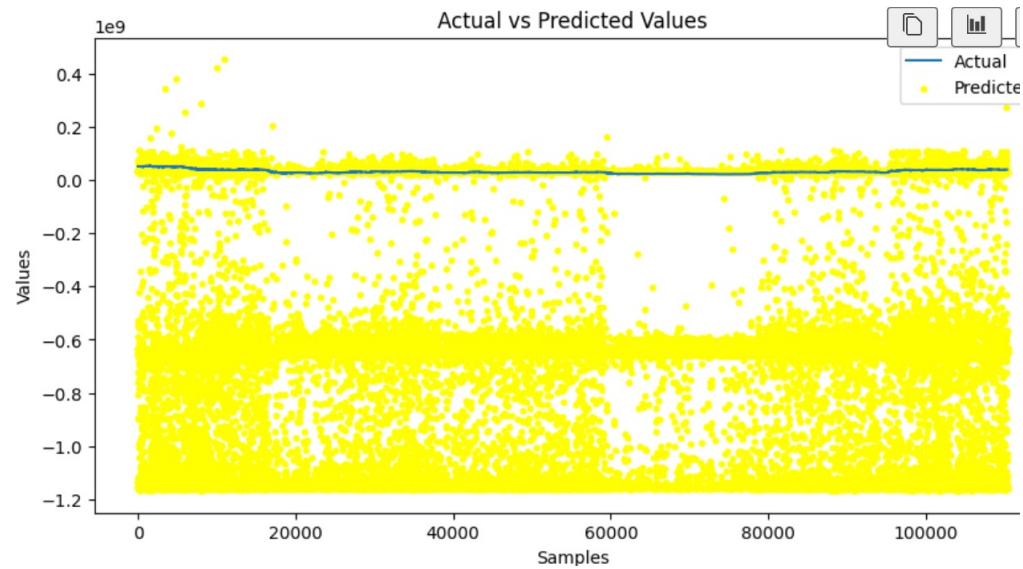
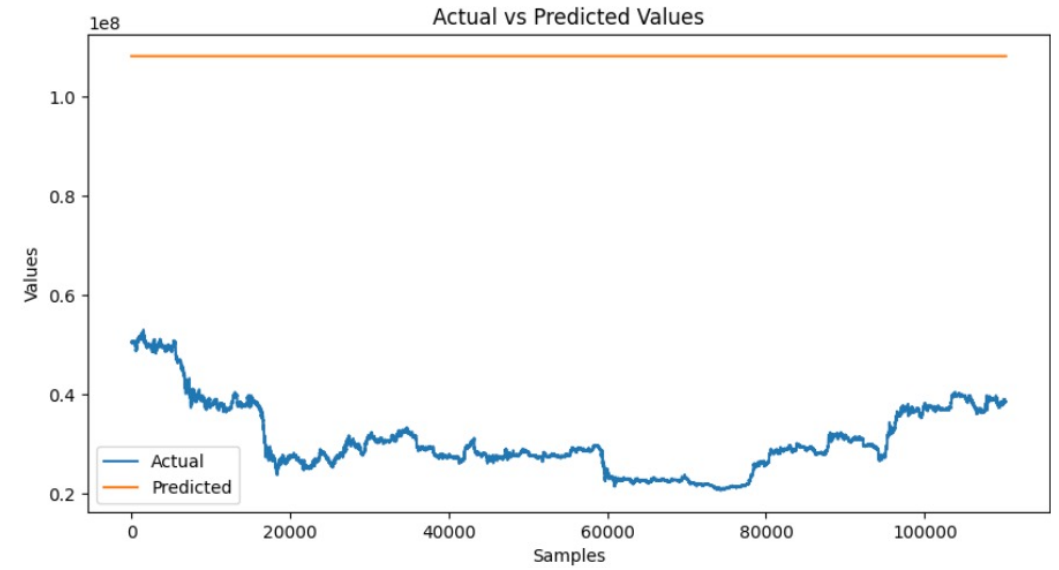
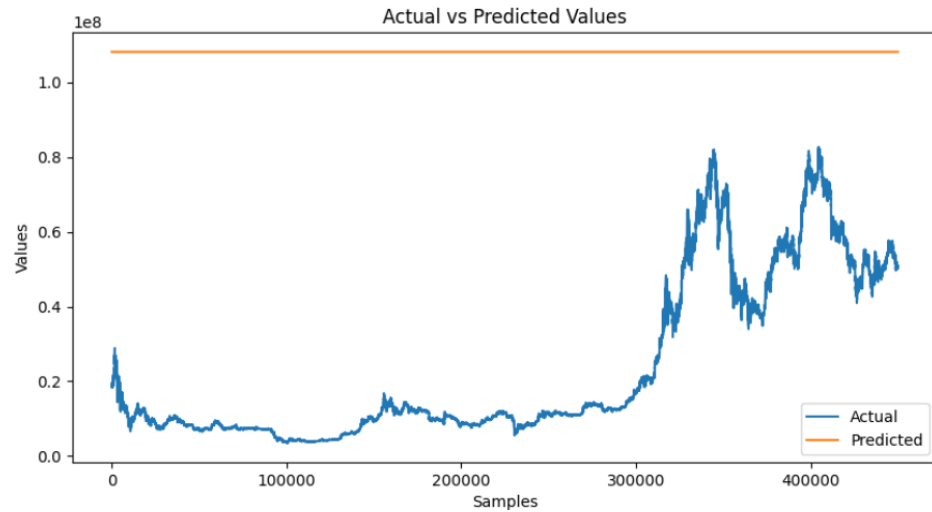
```
Epoch [2/20], Train Loss: 1023416922510546.8750, Val Loss: 977383097527322.1250
```

```
Epoch [3/20], Train Loss: 1013666488697633.5000, Val Loss: 964621839322575.2500
```

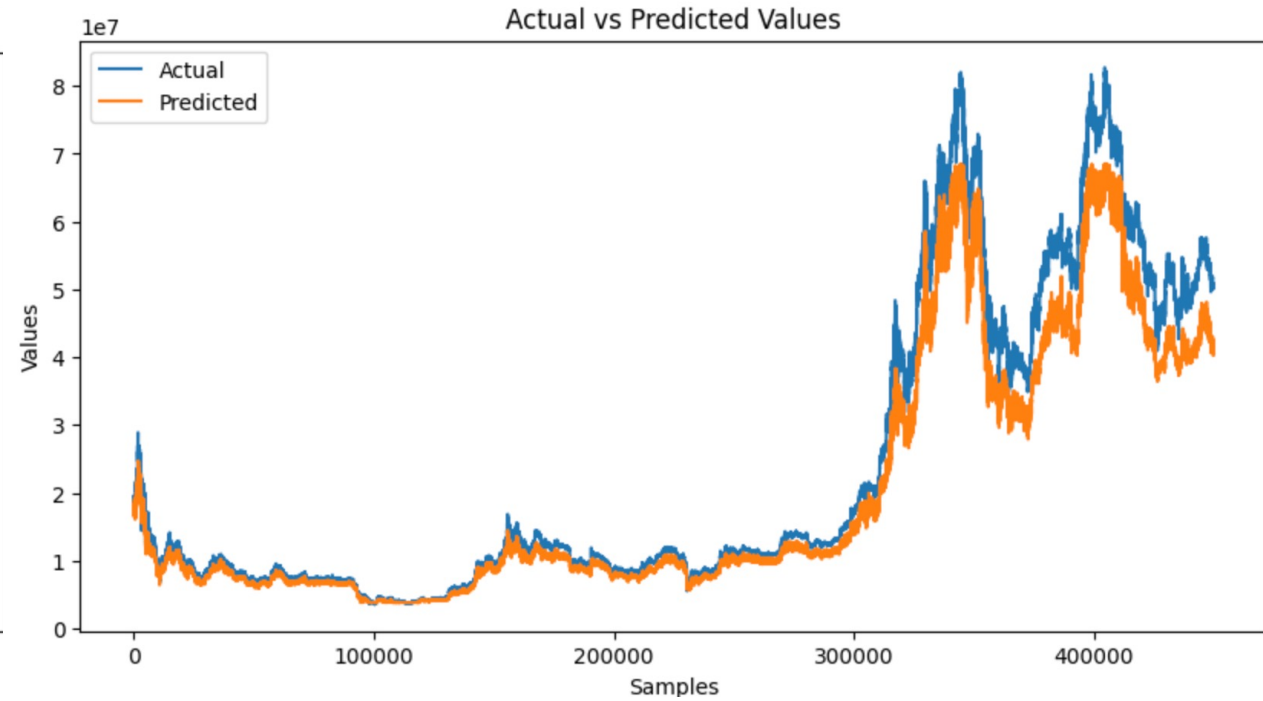
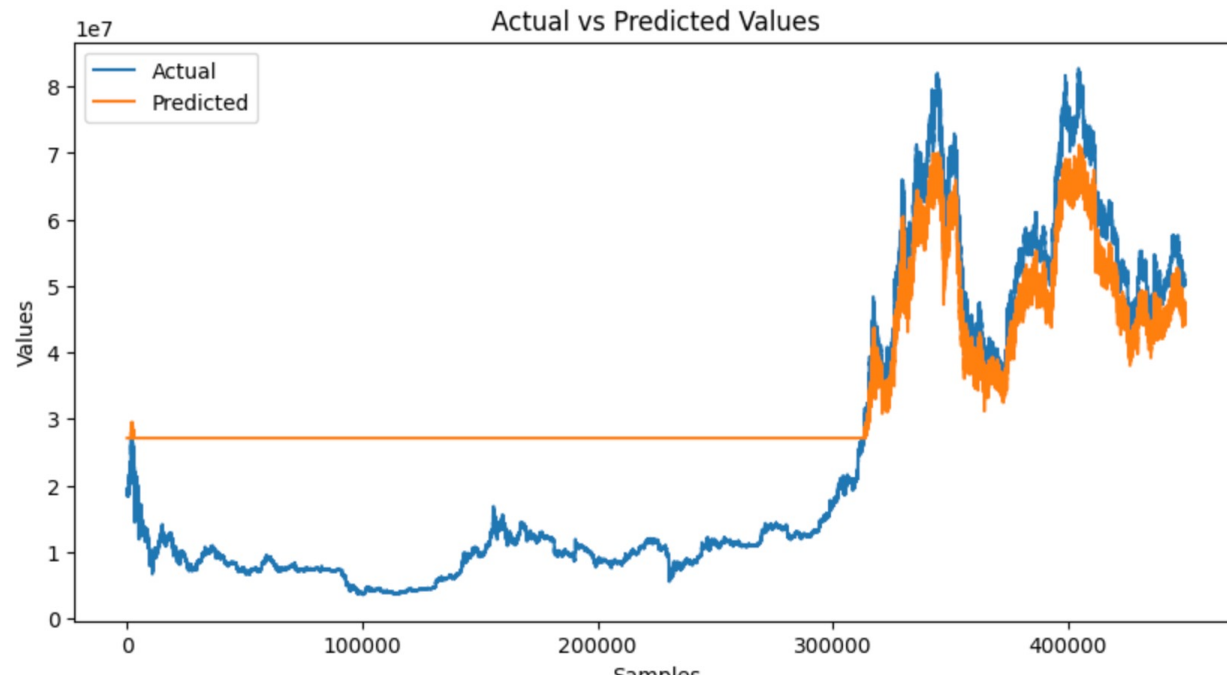
```
Epoch [4/20], Train Loss: 1004049198805061.8750, Val Loss: 951998523387780.5000
```

```
Epoch [5/20], Train Loss: 994550009695142.2500, Val Loss: 939511095389818.3750
```

Modeling error 2



Modeling error 3



QnA

