

삼성전자 주가예측 프로젝트

프로젝트 설명

딥러닝 모델을 이용하여 삼성전자
주가예측을 진행해보자

프로젝트 참가자

서강대학교 윤기웅

01 Index

프로젝트 동기	프로젝트 목표	데이터 수집/분석	사용 모델	모델 성능 평가	마치며
01 - 02	03- 05	06 - 09	10 - 14	15 - 26	27-28

01

Project Motivation/Goal

프로젝트 동기

- 코로나 이후 주식투자에 대한 사람들의 관심 증가
- 불안전해진 세계 경제, 주식시장의 변동성 증가

프로젝트의 목표

- 수급, 거래량을 이용한 주가예측 모델 개발하여 투자자들의 수익률 향상

01 --- Motivation



02 Motivation



동기

코로나 19 이후 국내 주식시장의 개인 투자자 비율은 코로나 19 이전 대비 약 70% 상승
20대의 수익률 특히, 20대 남성의 주식 수익률은 다른 연령대에 비해 현저히 낮음

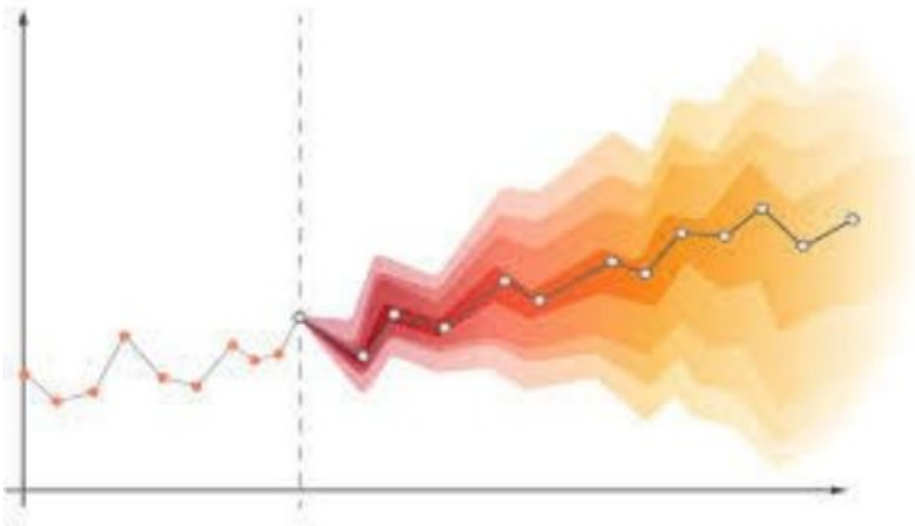


주식 시장의 변동성 증가, 외국인, 기관들의 거래에 영향을 많이 받기 때문



03 Goal

Stock Price Prediction with Deep Learning



시계열 데이터 예측을 위한 딥러닝 모델 이용

주가 흐름과 연관성 높은 코스피의 수급 데이터 활용



출처 : <https://medium.com/analytics-vidhya/time-series-forecasting-c73dec0b7533>

04 Goal



전체 코스피 시가총액의 20% 이상을 차지할 정도의 거대 기업
해당 종목 거래량, 코스피 수급 데이터를 활용하여 삼성전자 주가 예측



주가 예측 딥러닝 모델을 사용하여 주식 투자 수익률 향상



05 _____ Definitions

- Look Back window
- Look Ahead (Forecasting Horizon)
- Channel
- Close (종가)
- 순매수 (코스피 매수 금액 - 코스피 매도 금액)
- 거래량 (삼성전자 주식의 거래 총량)

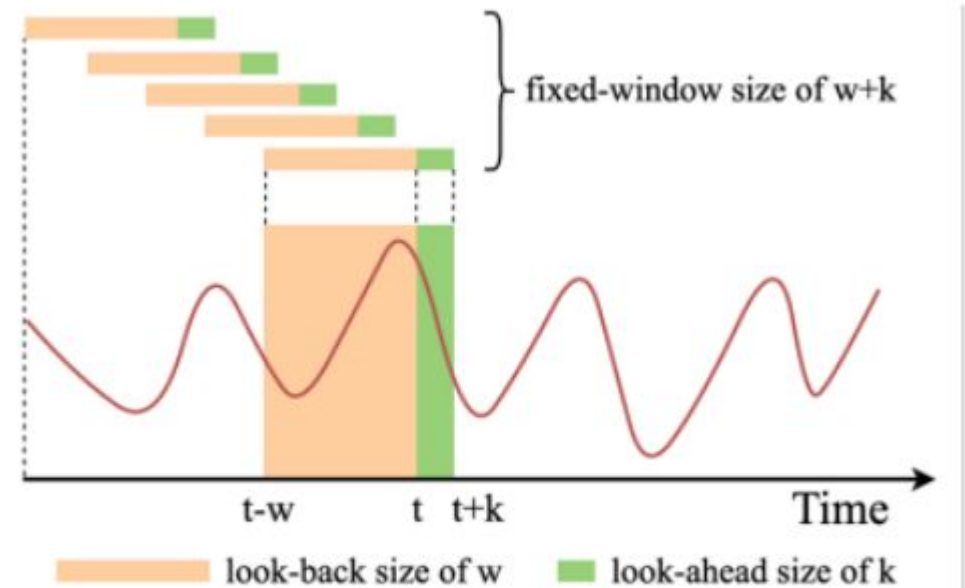


Fig. 1 Using a sliding window of width w to construct the training set for predicting k time steps in the future.

데이터 수집

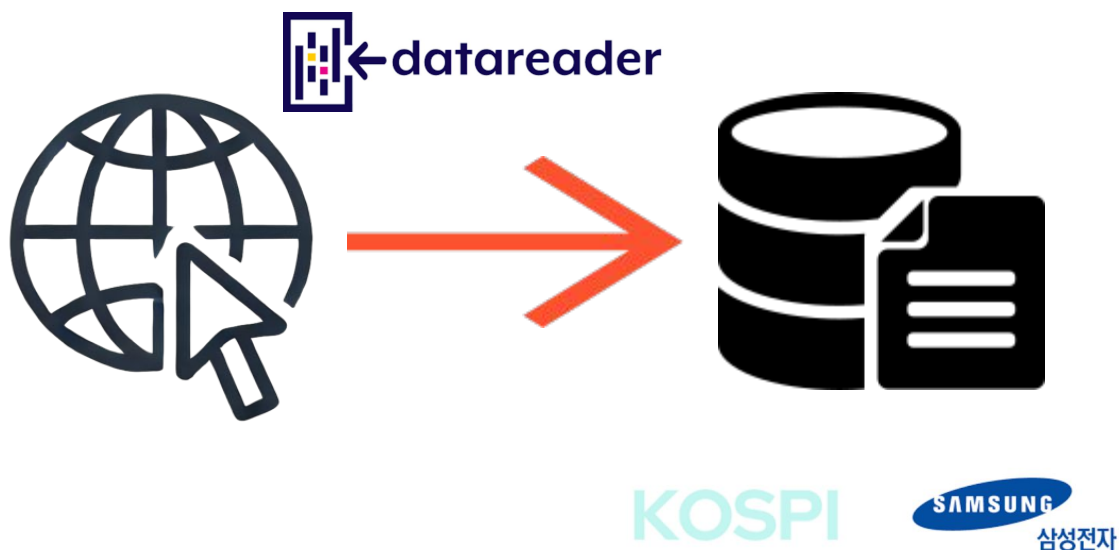
- pandas-datareader
 - web crawling
-

데이터 분석

02

Data Collect/Analysis

06 Data Collection



데이터 수집

- 방법 : pandas_datareader, web crawling
- 기간 : 2005년 1월 3일 ~ 2024년 6월 26일
- 항목 : 삼성전자 종가 및 일일 거래량, 코스피 개인, 외국인, 기관계, 기타법인의 순매수 데이터

출처 : https://blog.hectodata.co.kr/crawling_vs_scraping

https://pandas-datareader.readthedocs.io/en/latest/remote_data.html

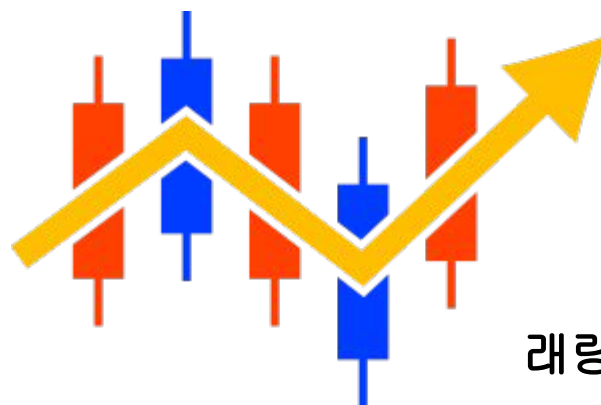
07 Data Collection



외국인 투자자



기관 투자자



래량



개인 투자자

거



08 Data Analysis

데이터 분석

- 거래량, 개인, 기관의 순매수 영향을 받는다
- 개인 순매수는 기관과 외국인 순매수와 상관성이 높다



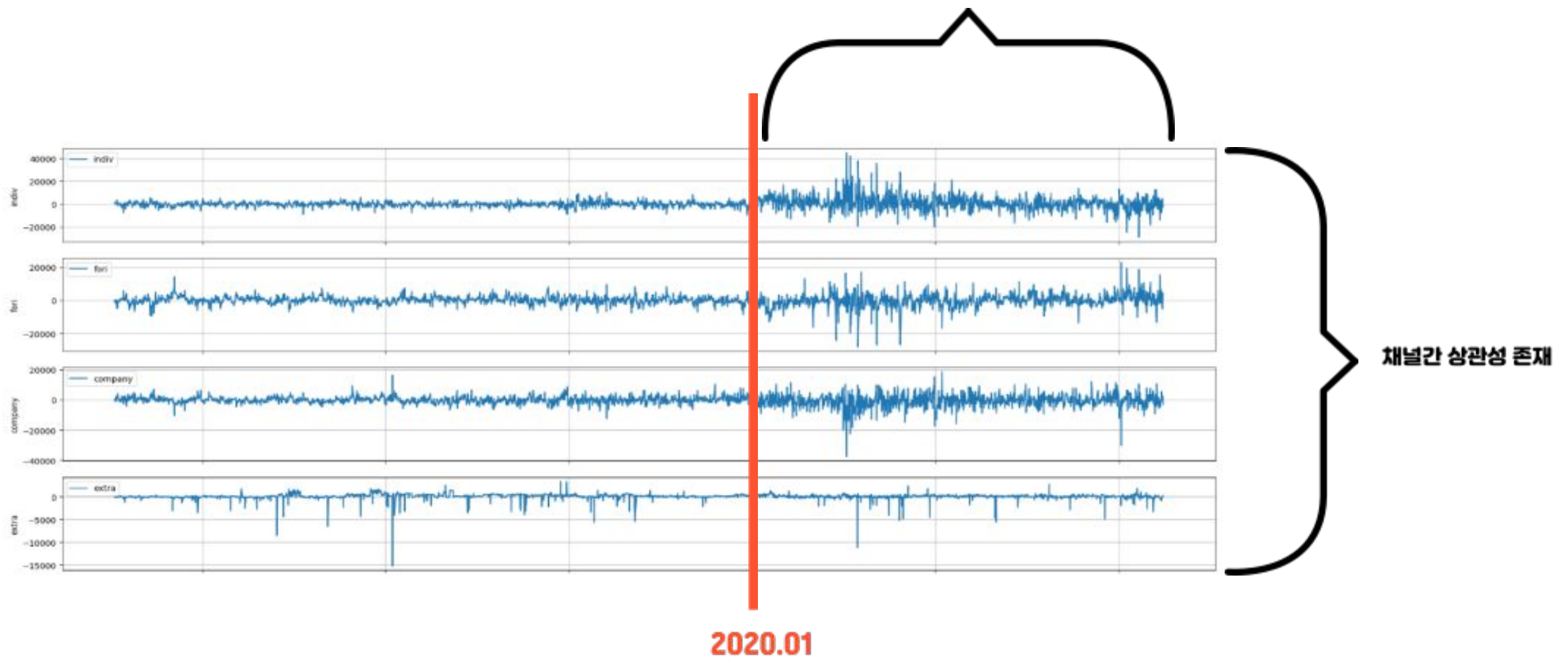
다변량 시계열 예측 시도하기

가격 탄력성을 새로운 특성으로 추가 해보기



09 Data Analysis

거래량, 변동성 상승



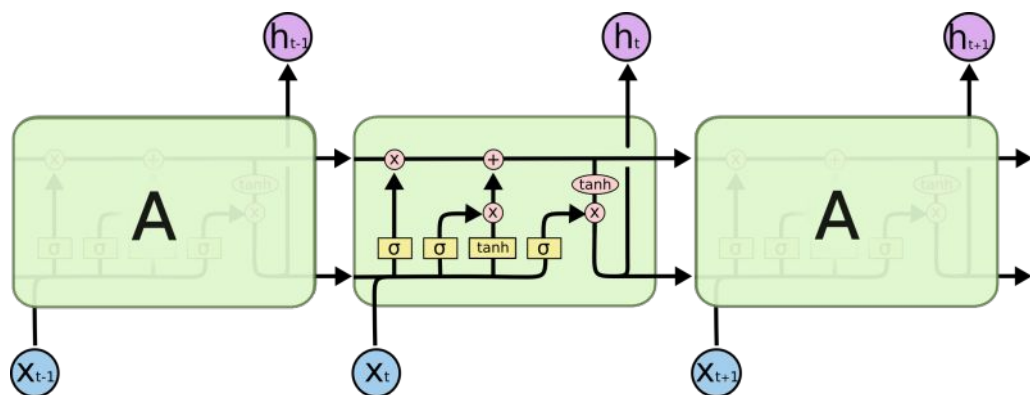
03

Models

-
- LSTM
 - LSTF-Linear(Dlinear)
 - Dlinear with Attention
 - Seg RNN
 - iTransformer

10

LSTM (Long Short-Term Memory)

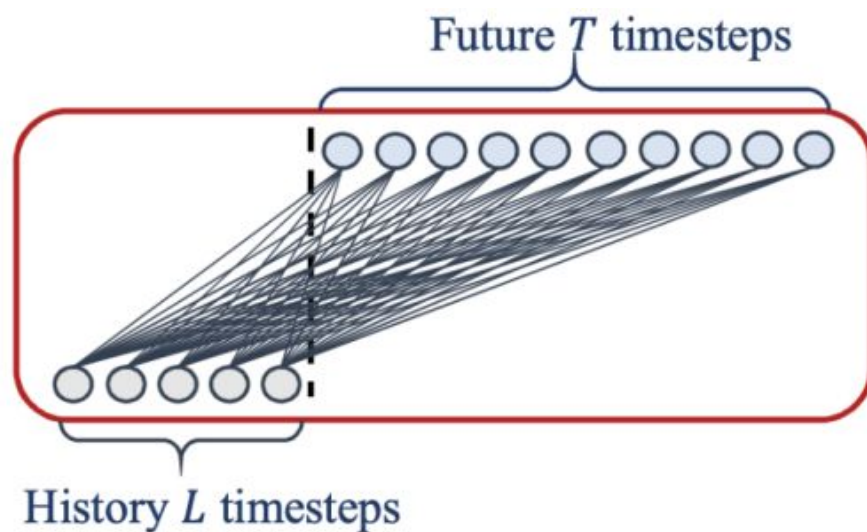


Model Characteristics

- 입,출력 게이트, 망각 게이트를 이용한 RNN의 단점 개선
- 최신 모델들에 비해 아쉬운 성능

11

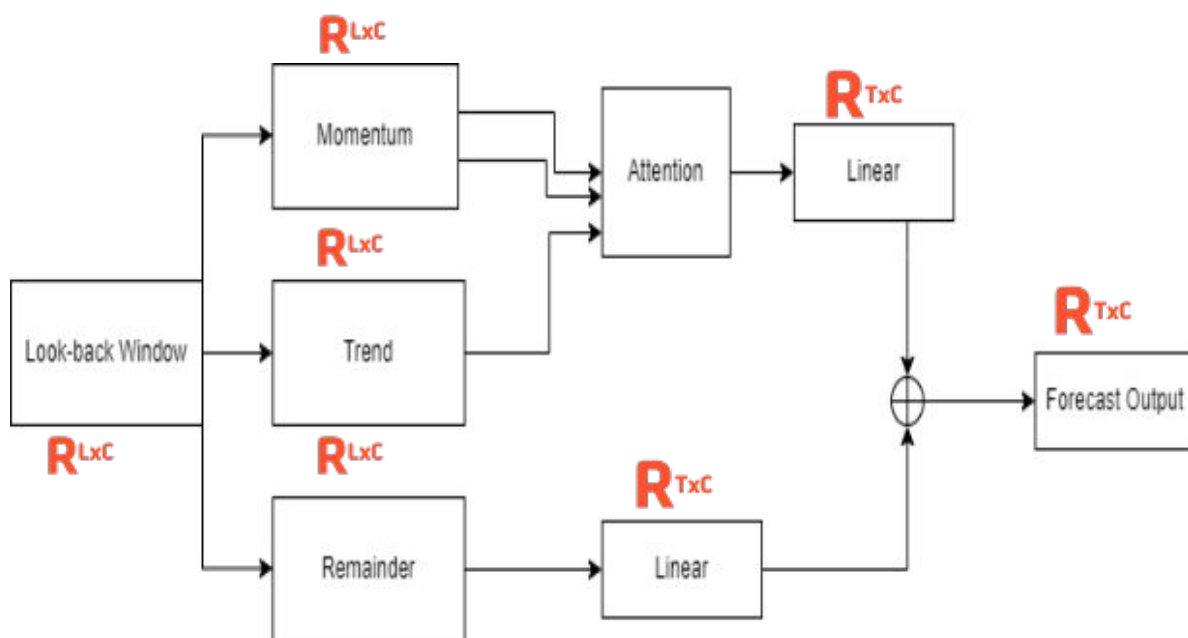
LSTF-Linear(Dlinear)



Model Characteristics

- 단층 선형 모델로 기존 Transformer 모델의 성능 능가
- 트렌드 요소, 계절성 요소를 분해 하여 예측 실행
- 트렌드 요소가 명확해야 모델 성능이 높아지는 한계점 존재

12 Dlinear with Attention



Model Characteristics

- 기존 Dlinear에 더해 **모멘텀 요소**까지 사용
- 어텐션 메커니즘을 이용한 채널 간 관계 파악



기존 Dlinear 모델에 비해 더 낮은 loss
주가 트렌드, 변동에 대한 예측 능력 향상

13 SegRNN

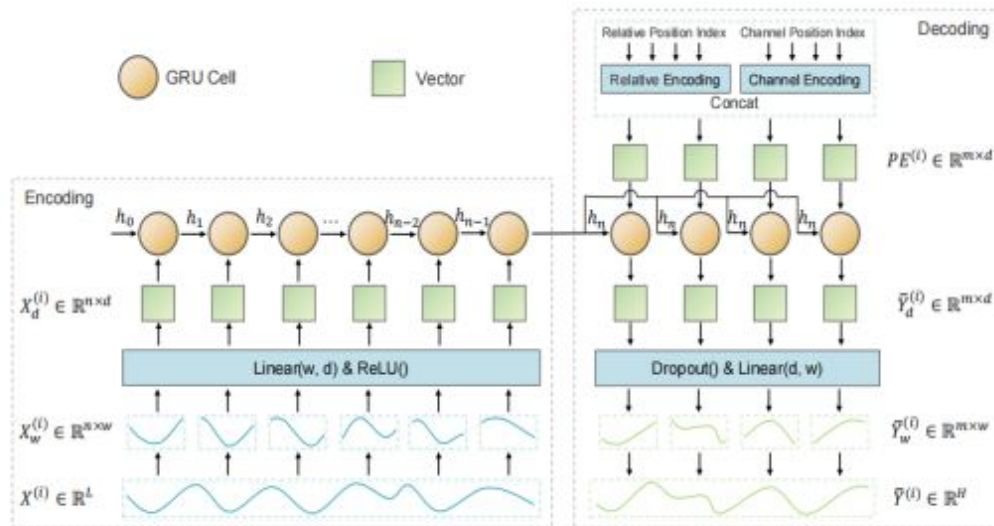


Figure 4: The model architecture of SegRNN.

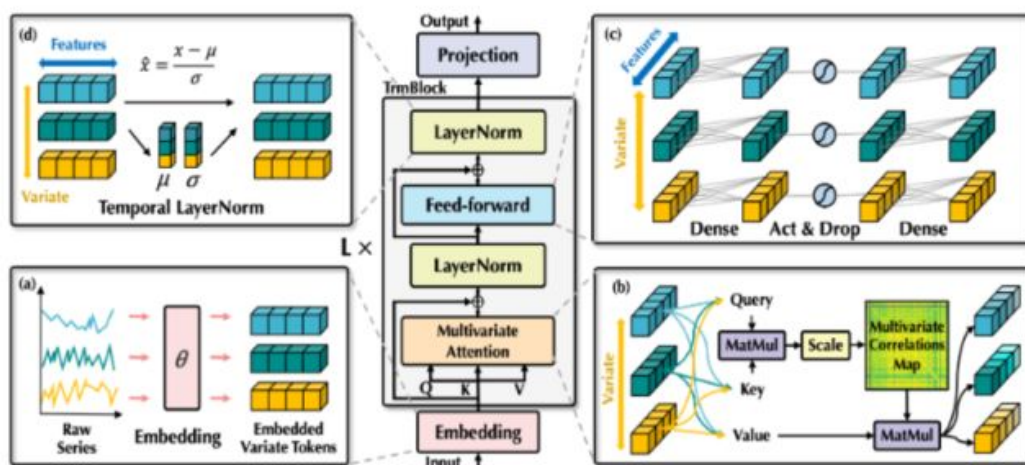
Model Characteristics

- Segment wise iterations 방법 채택
- Parallel Multistep Forecasting 방법으로 성능 향상



Multivariate forecasting benchmarks 에 대한 뛰어난 성능

14 iTransformer



Model Characteristics

- 개별 채널들을 Token으로 간주, 각각에 대한 임베딩 생성
- 데이터 상관성 학습을 위한 Attention 매카니즘 사용



Multivariate forecasting benchmarks 에 대한 뛰어난 성능

04

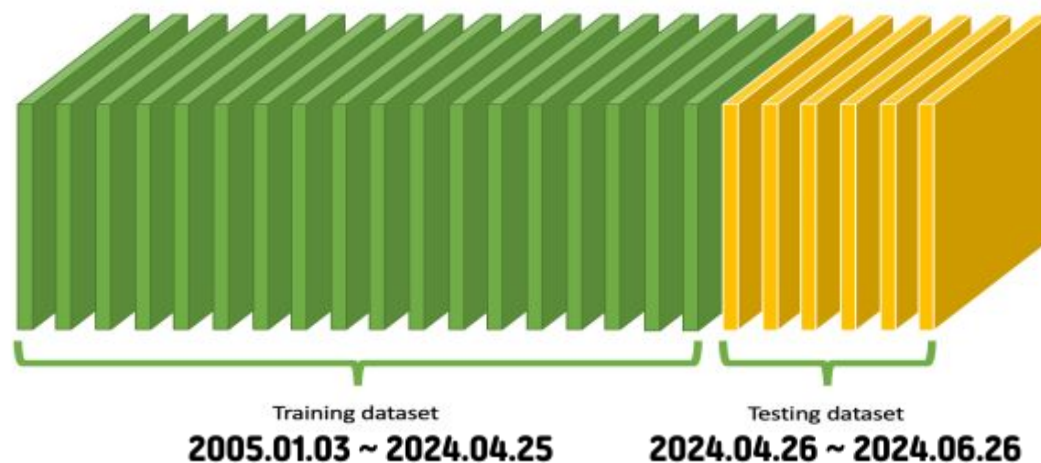
Results

Model Evaluation

Model Performance

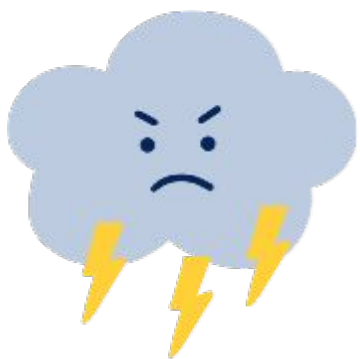
15 Model Evaluation

Train/Test Split



$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

16 _____ Model Performance



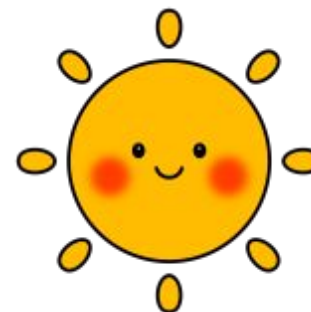
LSTM



Dlinear



DlinearwithAttention



iTransformer



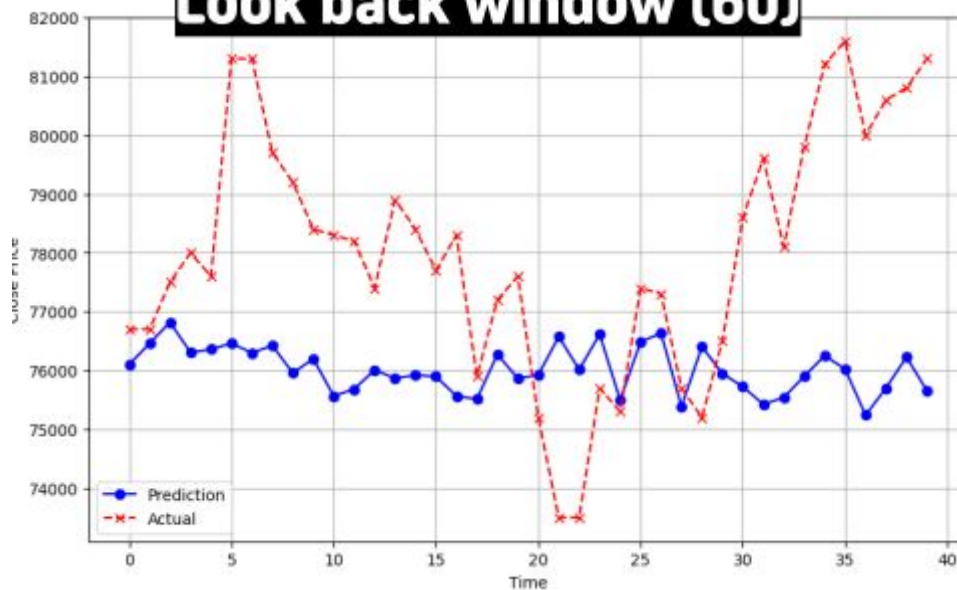
SegRNN

17 **Model Performance**

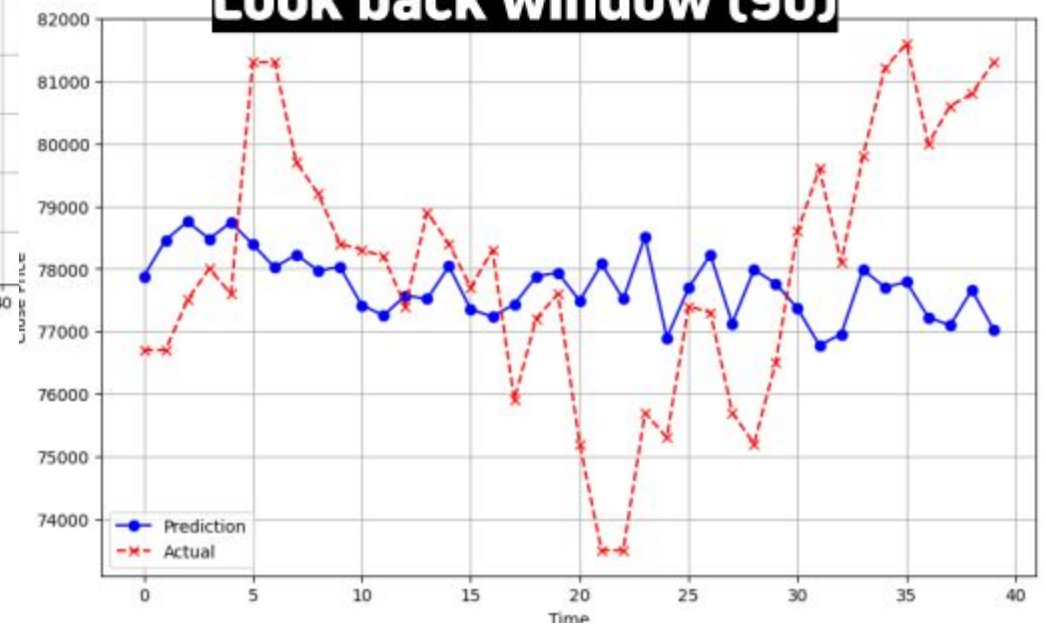
Model Look back window	LSTM	Dlinear	DlinearwithAttention	SegRNN	iTransformer
60	2951	2711	2391	2372	2498
90	2194	2728	2220	1985	2017
120	2419	2576	2142	2139	2491
150	3320	2669	2431	2164	2168

18 LSTM

Look back window (60)

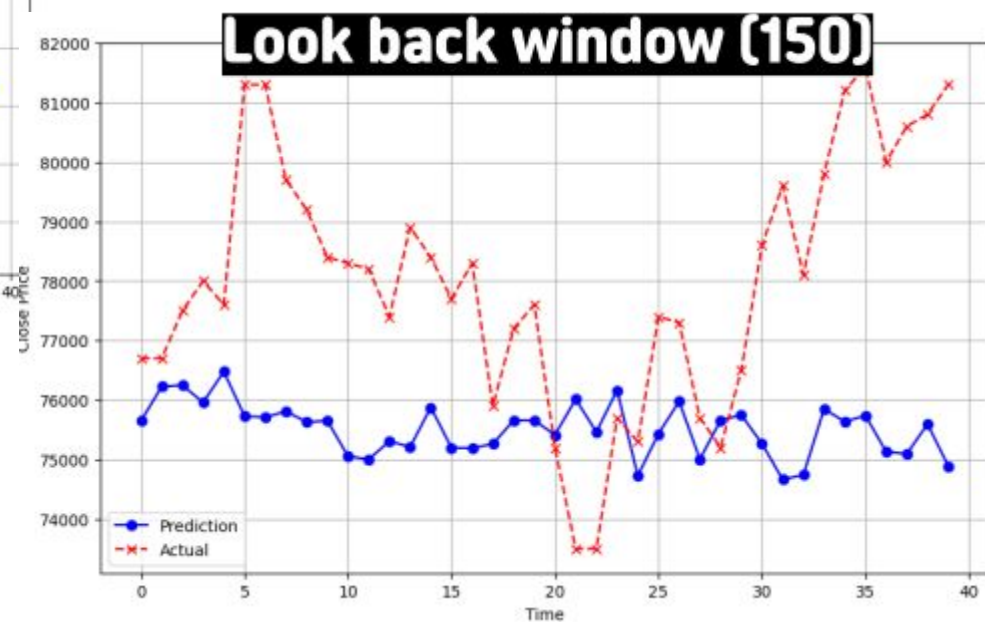
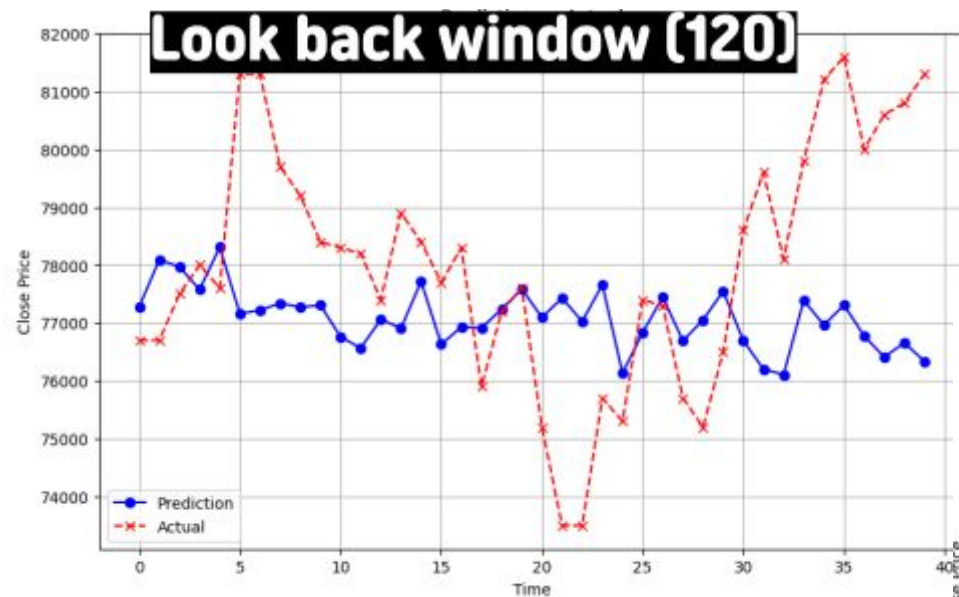


Look back window (90)

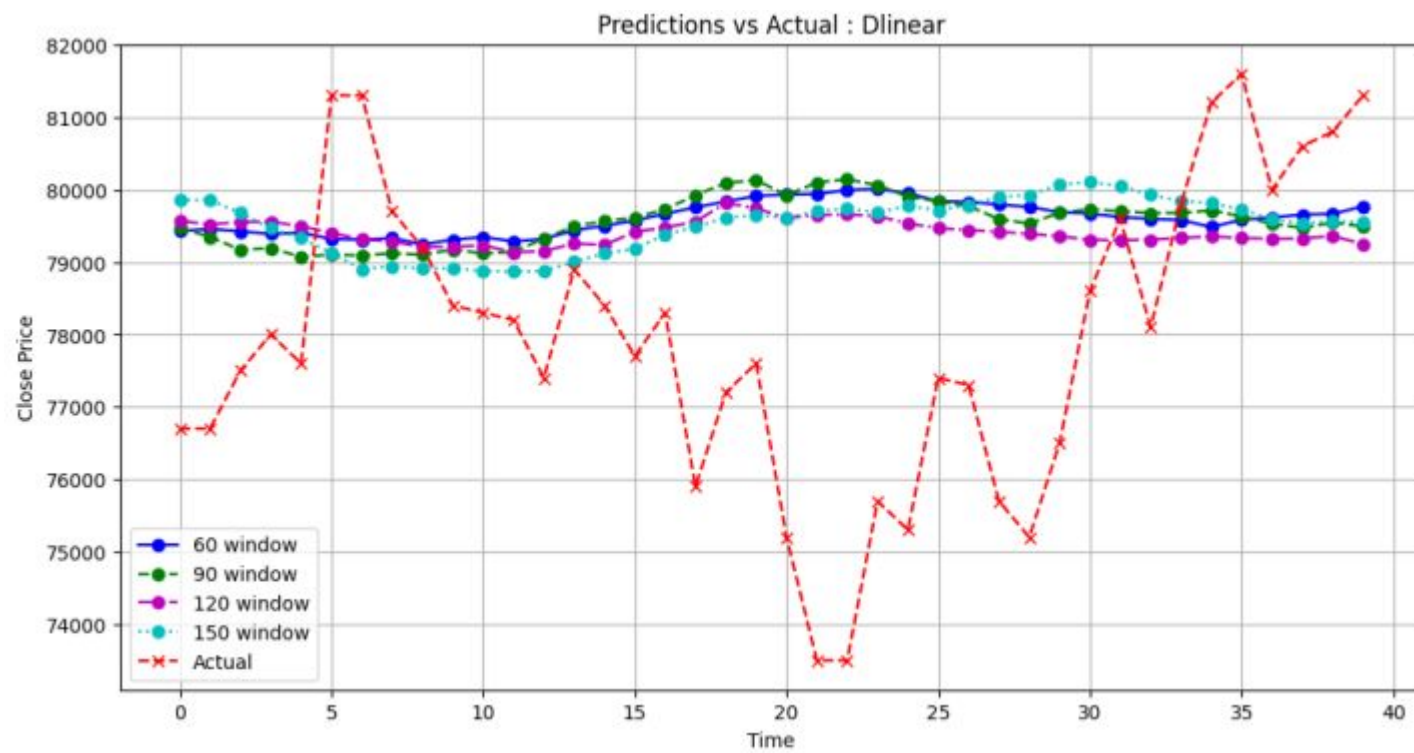


19

LSTM



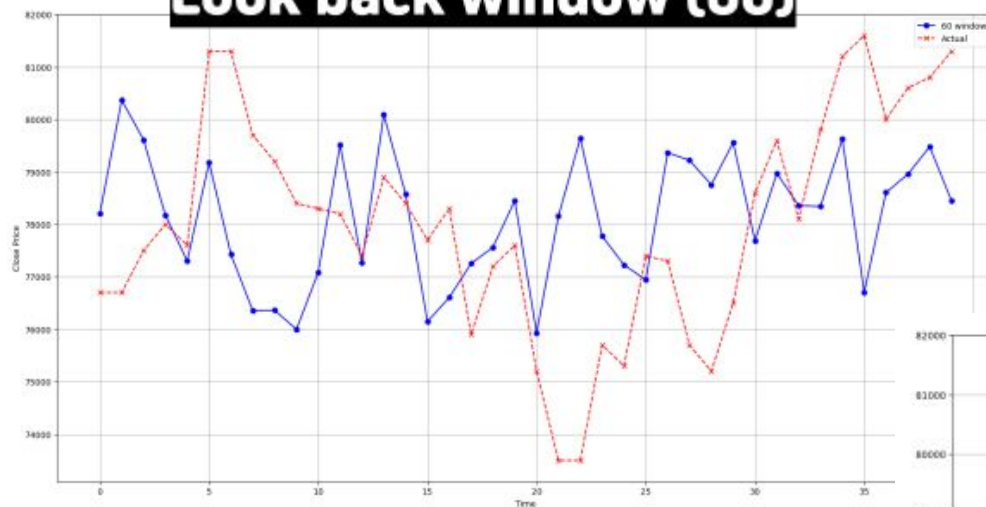
20 Dlinear



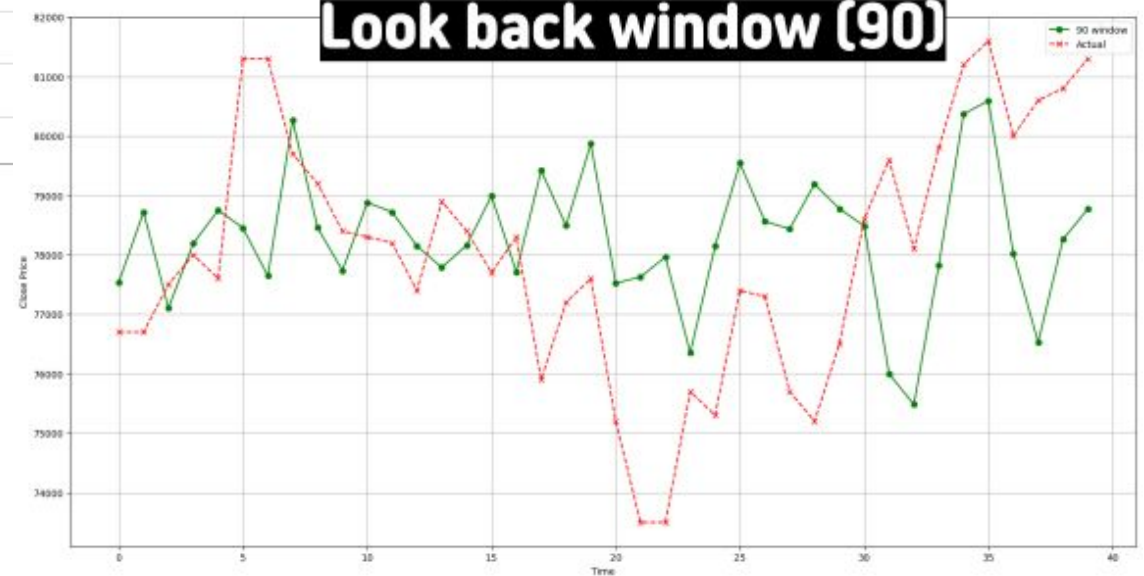
21

DlinearwithAttention

Look back window (60)



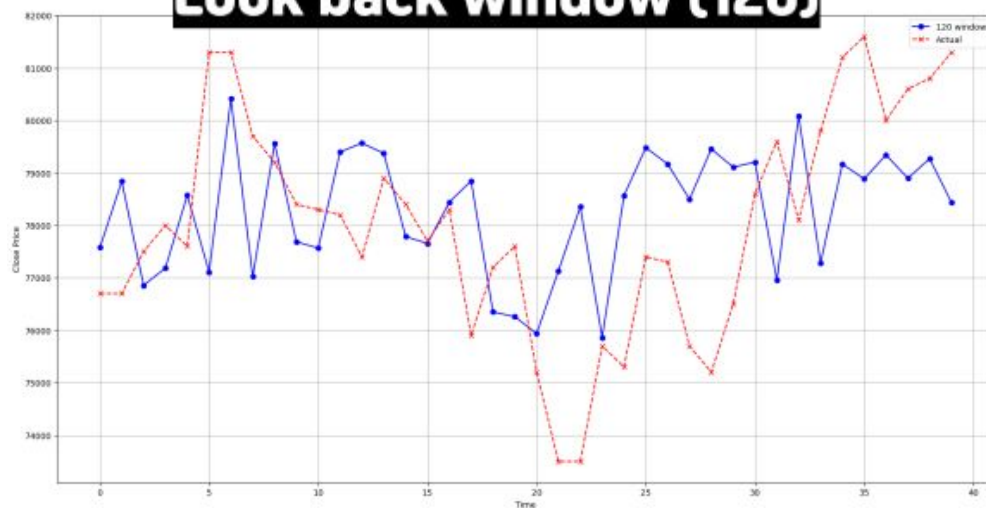
Look back window (90)



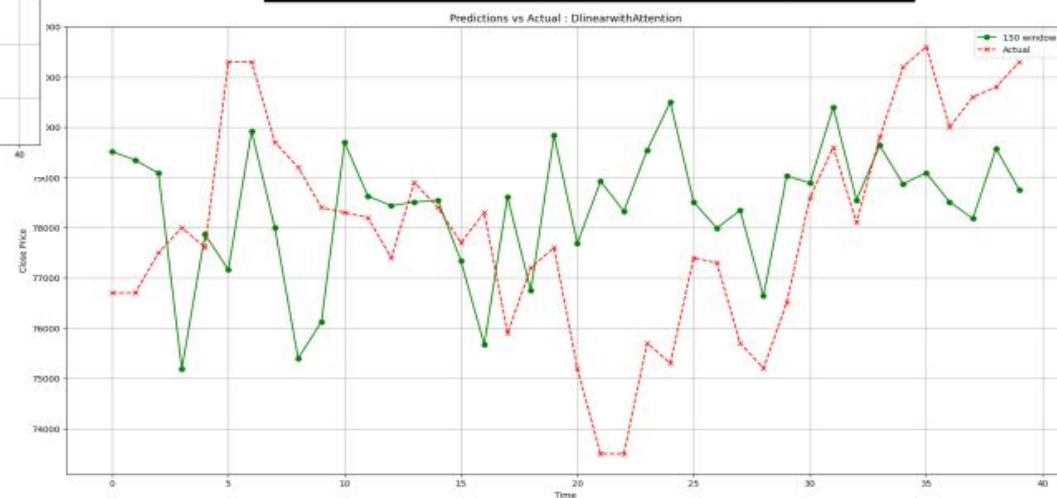
22

DlinearwithAttention

Look back window (120)

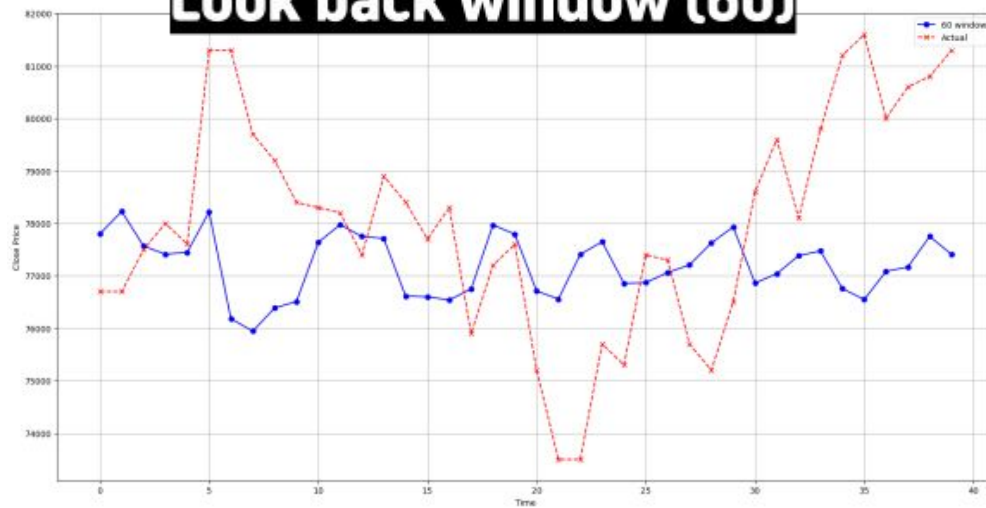


Look back window (150)

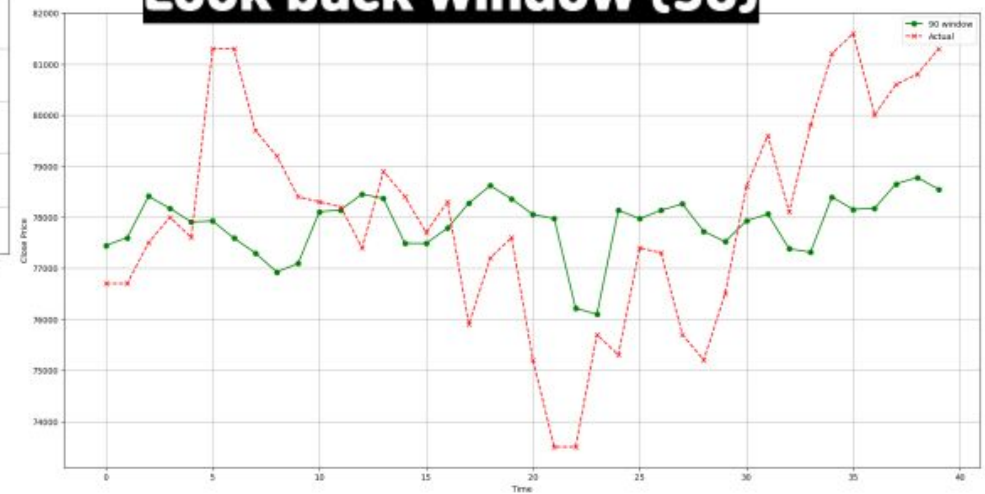


23 SegRNN

Look back window (60)

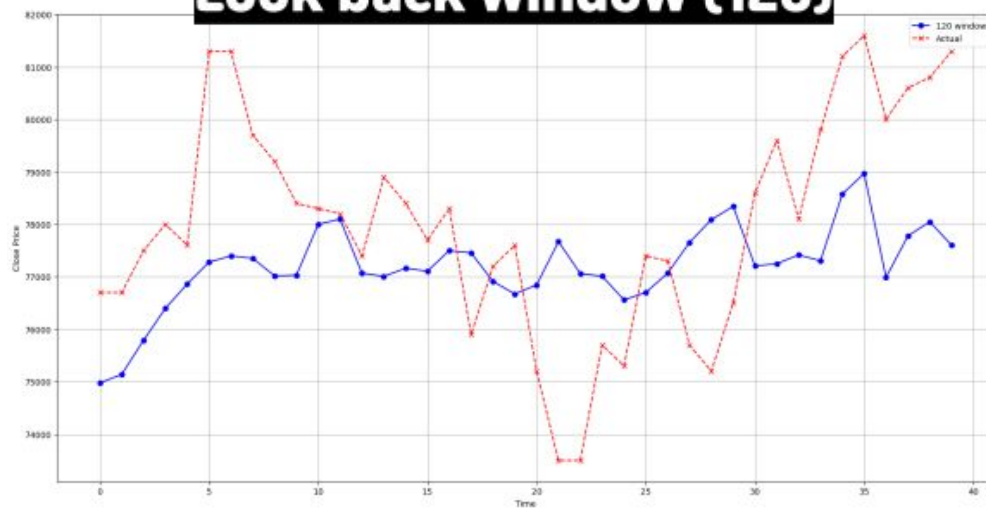


Look back window (90)

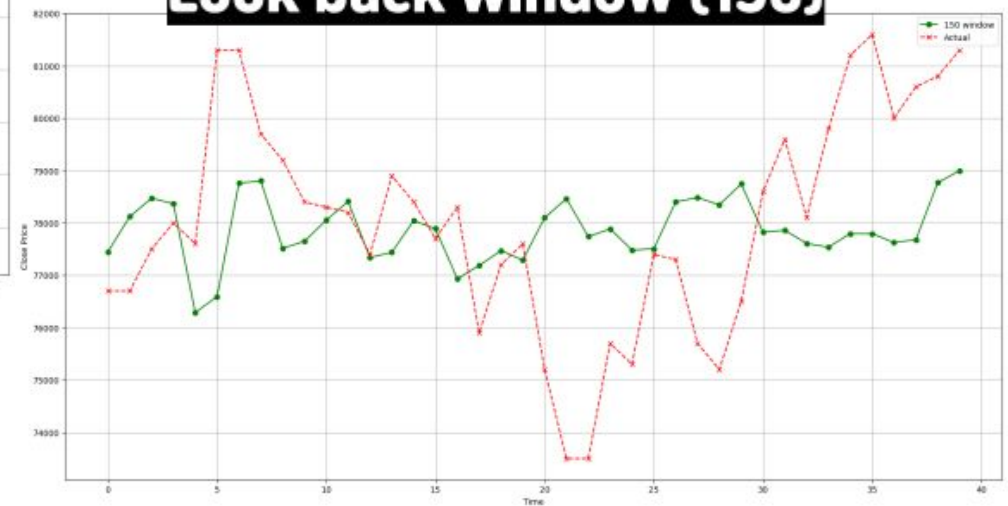


24 SegRNN

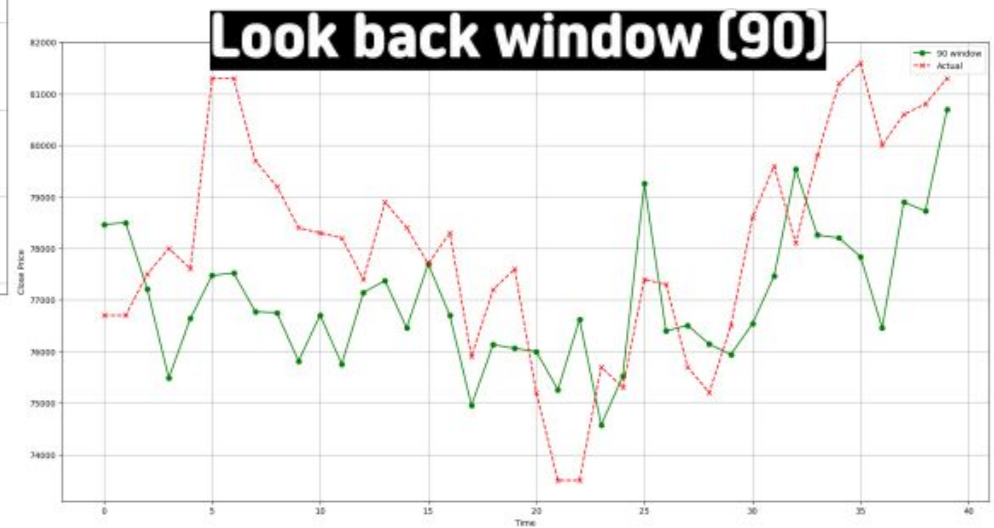
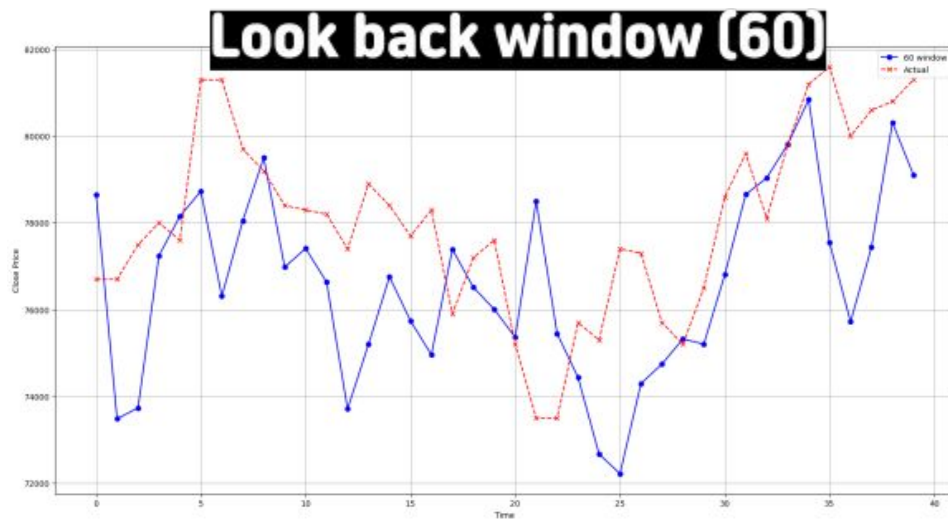
Look back window (120)



Look back window (150)

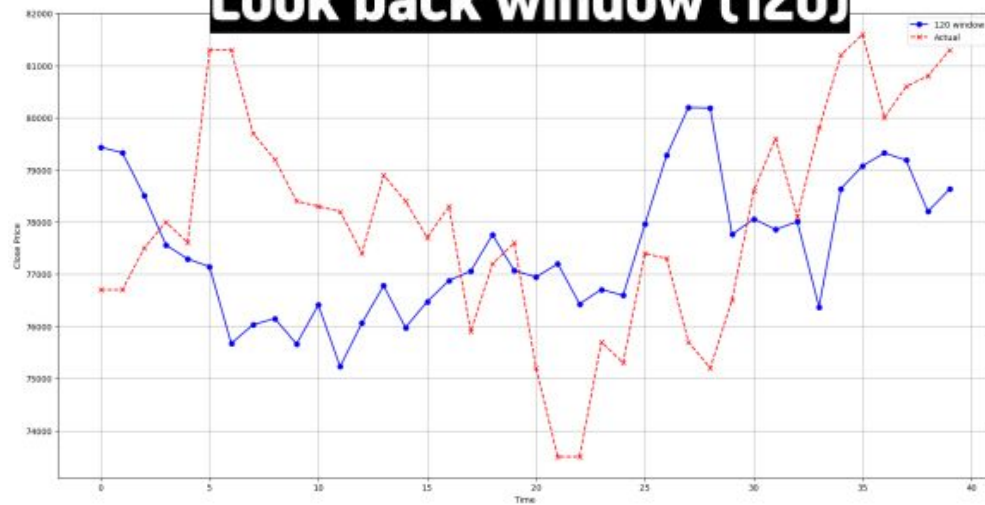


25 iTransformer

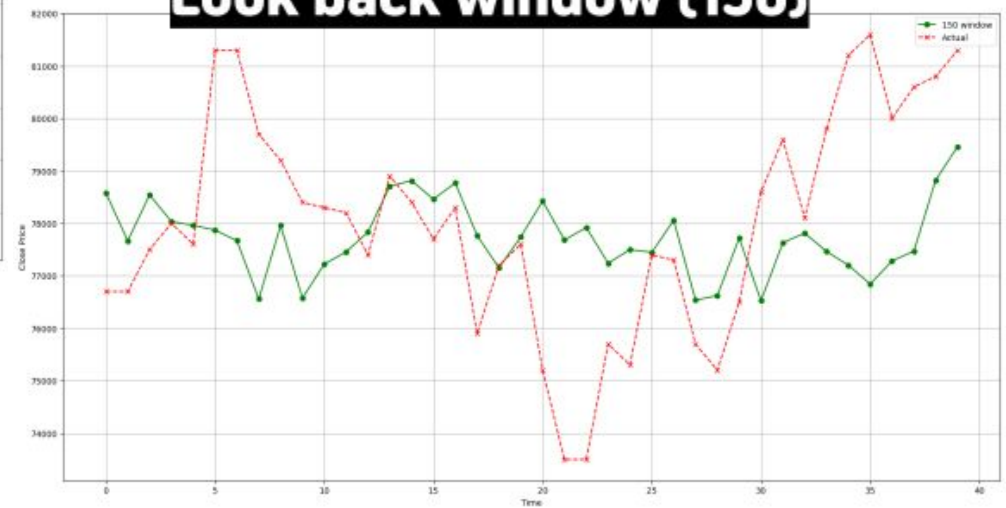


26 iTransformer

Look back window (120)



Look back window (150)



05

**Future Improvements/
Reflections**

Future Improvements

Reflections

27 **Future Improvements**



- 경쟁 기업의 주가 데이터 활용
- 해당 기업의 순매수 데이터 활용

-
- 다변량 시계열 데이터에 적합한 메커니즘 탐색
 - custom loss function 적용
 - 최적 하이퍼파라미터 탐색

28 Reflections



여러 변수의 영향을 받는 주가를 예측하기란 어렵다
하지만, 주가 흐름과 추세를 예측하는 것은 가능하다

29
End

♥ 감사하합니다^^ ♥

윤기웅의 발표였습니다!
들어주셔서 감사합니다