Section 2 Project

AI 부트캠프 1기 - 노태윤

1. Data (1) - Tesla Stock Price Dataset

- **Why Stock Price?**
  - 앞으로 도움이 될 것으로 생각함
- **Why Tesla?**
  - 배터리 & 전기차 분야에 관심이 많음
- **Data Size**
  - 2579 X 7

날짜	시가	최고가	최저가	종가	조정 종가	거래량

	Date	Open	High	Low	Close	Adj Close	Volume
0	2010-06-30	5.158	6.084	4.660	4.766	4.766	85935500
1	2010-07-01	5.000	5.184	4.054	4.392	4.392	41094000
2	2010-07-02	4.600	4.620	3.742	3.840	3.840	25699000
3	2010-07-06	4.000	4.000	3.166	3.222	3.222	34334500
4	2010-07-07	3.280	3.326	2.996	3.160	3.160	34608500

1. Data (2) - Modeling & Evaluation Metrics

- **Modeling**
  - Regression Model
  - Ensemble Model (Random Forest)

- **Evaluation Metrics**
  - MAE (Mean Absolute Error)
  - MSE (Mean Squared Error)
  - RMSE (Root Mean Squared Error)
  - R^2 (Coefficient of Determination)

1. Data(3) - Target (2010 ~ 2020)

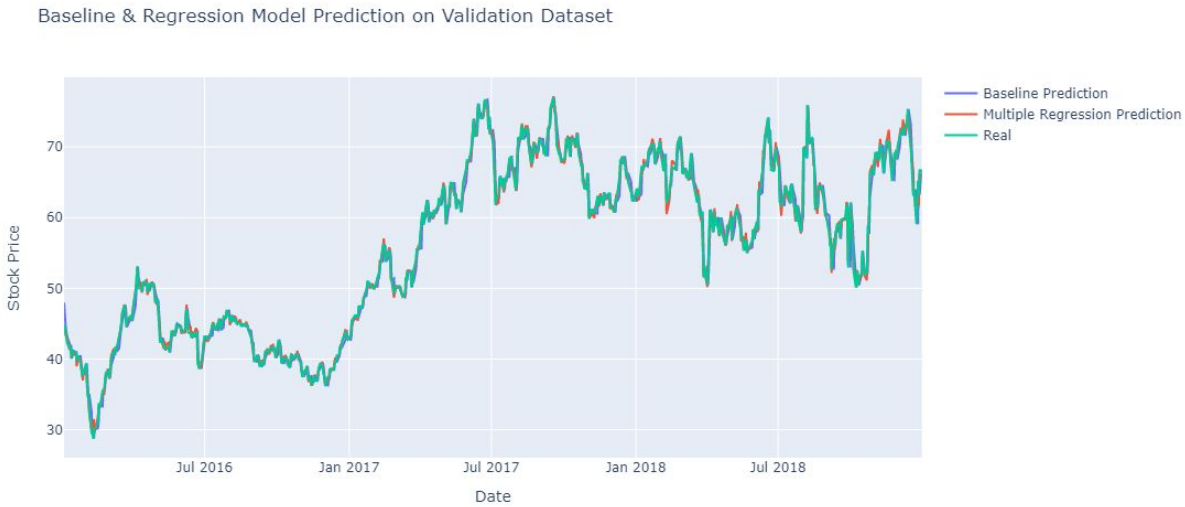
2. Simple Modeling(1) - Hypothesis & Baseline Model

- **Hypothesis**
  - Open 증가 → Close 증가
  - Open 감소 → Close 감소

- **Baseline Model**
  - 오늘 예측 종가 = 전일 종가

- **Data Split**
  - Train : 2010 ~ 2016
  - Validation : 2017 ~ 2018
  - Test : 2019 ~ 2020

2. Simple Modeling (2) - Model Analysis



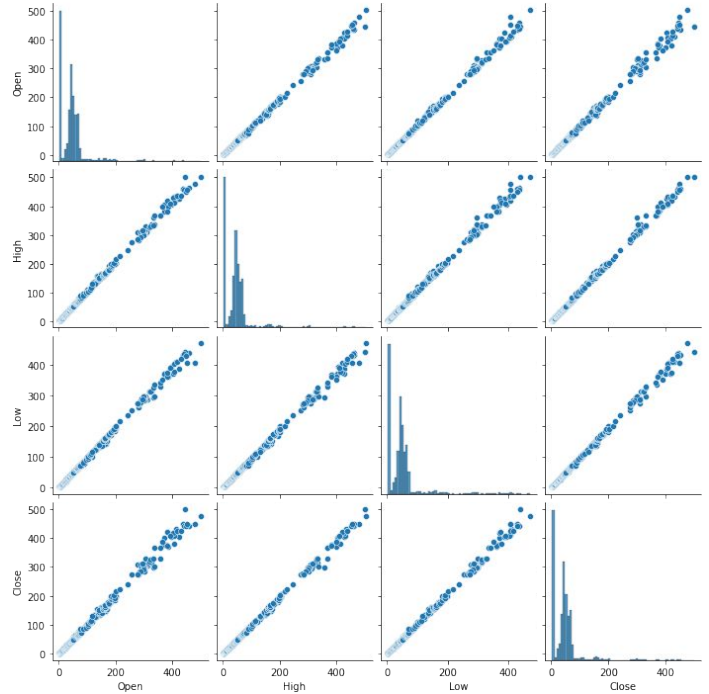
- **Baseline Model**
  - MSE : 2.67
  - MAE : 1.13
  - RMSE : 1.63
  - R^2 : 0.98

- **Multiple Regression Model**
  - MSE : 0.30
  - MAE : 0.42
  - RMSE : 0.55
  - R^2 : 0.99

2. Simple Modeling (3) - Limitation

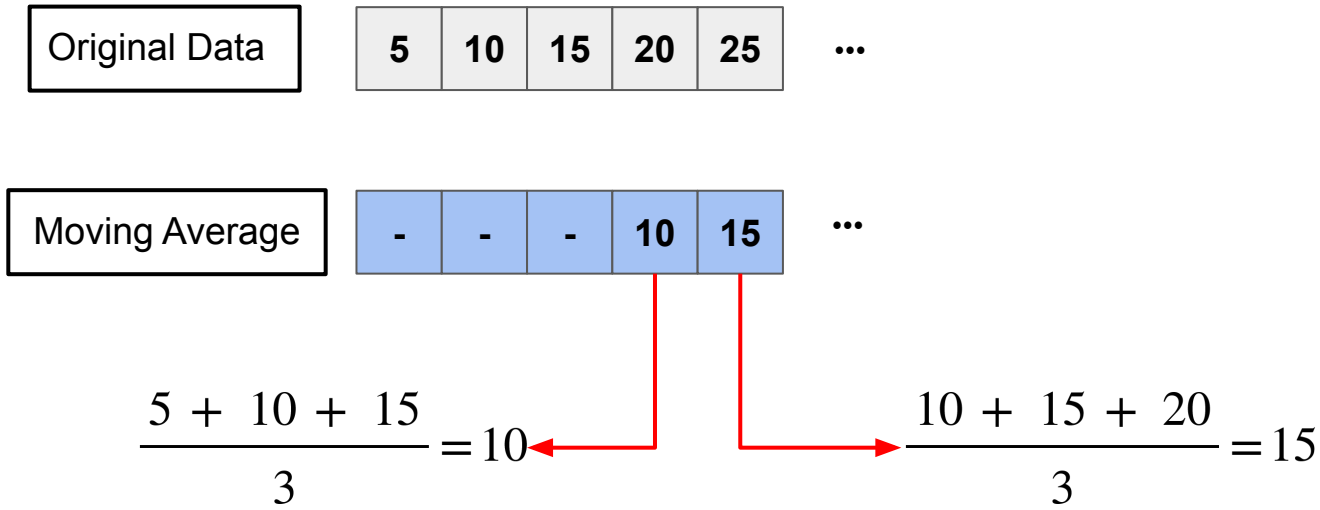
- **Limitation**
  - Target Leakage (High Correlation)
  - High & Low 정보를 사전에 알 수 없음

- **Improvements**
  - Open을 기준으로 한 모델



3. Moving Average (1) - Example

- 3 Day Moving Average

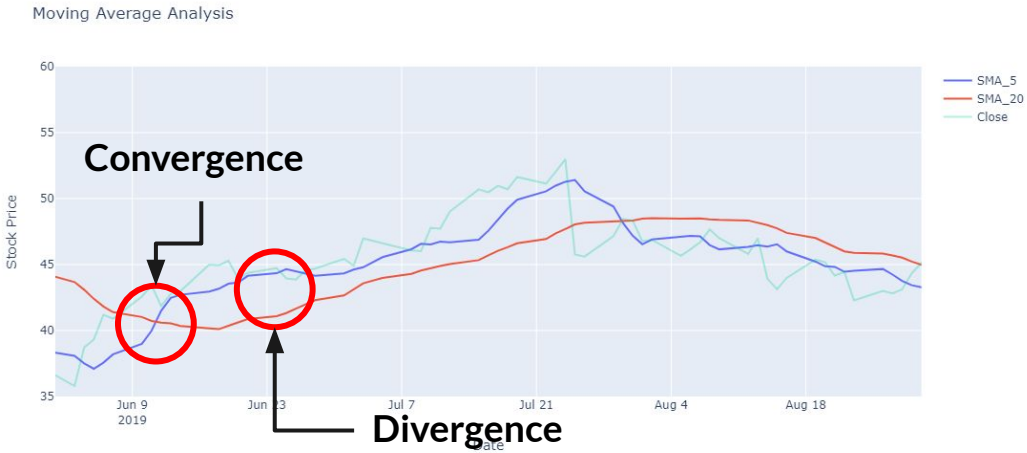


3. Moving Average (2) - Example

3. Moving Average (3)

- Simple Moving Average

- Exponential Moving Average
  - 현재 데이터에 큰 가중치
  - 과거 데이터에 적은 가중치



- Moving Average Convergence Divergence

4. New Model (1)

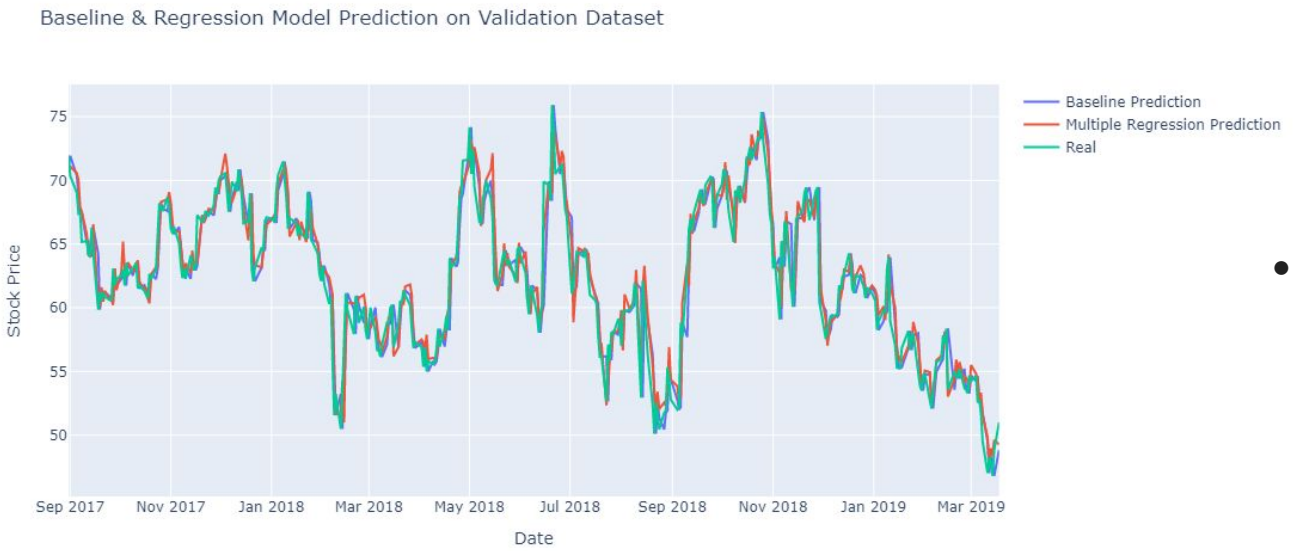
- Data Wrangling
  - Nasdaq 종합 지수 (Tesla 2010 년6월 Nasdaq 상장)
  - Tesla 주식 데이터 업데이트 (2월 15일 까지)

- 시가 feature engineering
  - 시가 이동 평균 법 적용
  - 시가 등락률 & 변동성 추가

- 시가, 최저가, 최고가, 거래량 제거

- Data Split
  - 75% / 15% / 15% (시간 순서)

4. New Model (2) - Regression Model



- **New Baseline Model**
  - MSE : 4.30
  - MAE : 1.47
  - RMSE : 2.07
  - R^2 : 0.87

- **New Regression Model**
  - MSE : 2.81
  - MAE : 1.25
  - RMSE : 1.67
  - R^2 : 0.91

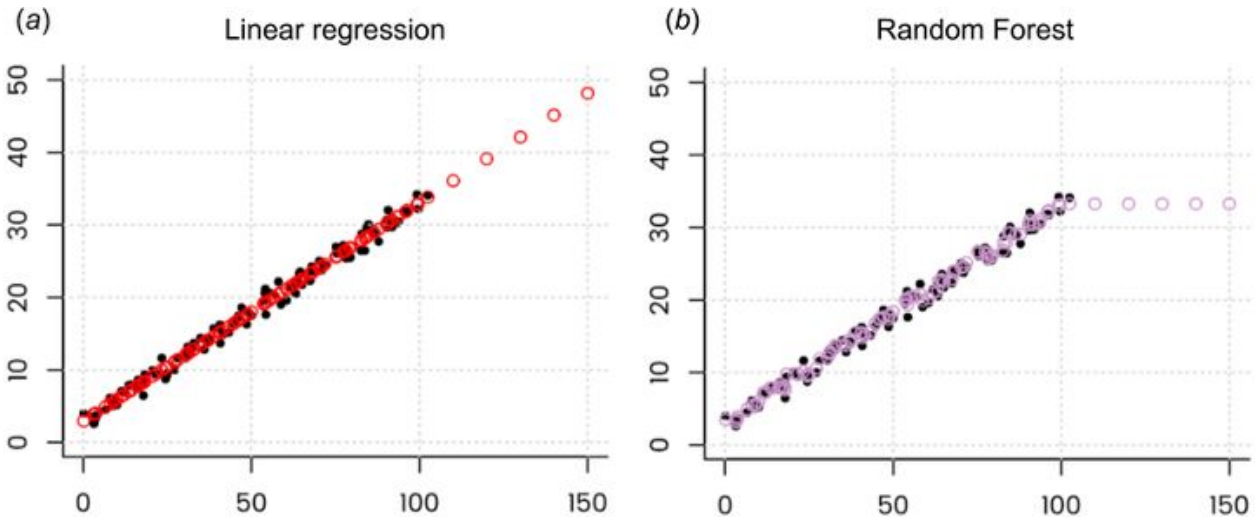
4. New Model (3) - Ensemble Model

- **Random Forest Regressor Model**
  - MSE : 20.58
  - MAE : 3.69
  - RMSE : 4.53
  - R^2 : 0.39

- **XGB Regressor Model**
  - MSE : 21.34
  - MAE : 3.88
  - RMSE : 4.61
  - R^2 : 0.36

4. New Model (4)

- Random Forest Model
  - Extrapolation (X)



4. New Model (5) - Hyperparameter Tuning

- **New Regression Model**
  - MSE : 2.814
  - MAE : 1.250
  - RMSE : 1.677
  - R^2 : 0.916

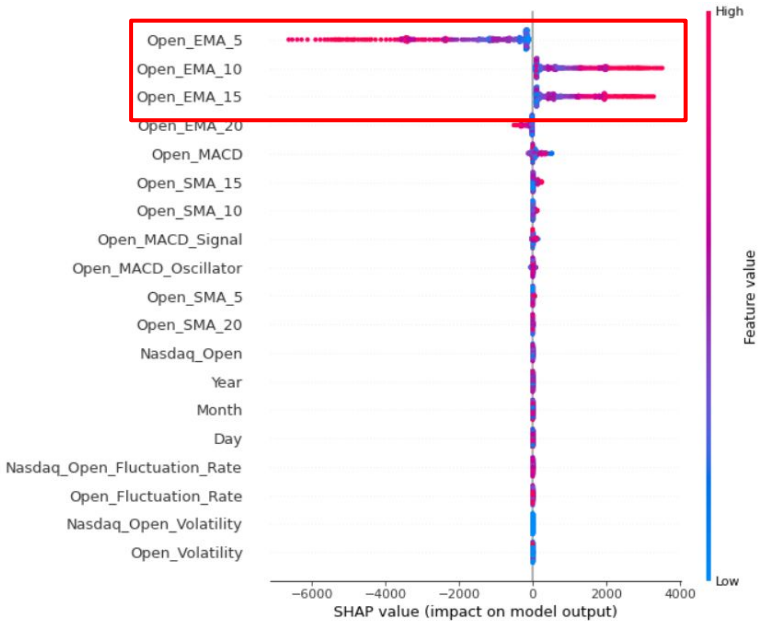
- **Ridge Regression Model**
  - MSE : 2.795
  - MAE : 1.241
  - RMSE : 1.671
  - R^2 : 0.917

5. New Model (6) - Test Data

- **Ridge Regression Model**
  - MSE : 125.51
  - MAE : 6.23
  - RMSE : 11.20
  - R^2 : 0.99

6. Model Analysis - Permutation Importance & Shapley Values

Weight	Feature
143.7187 ± 14.1671	Open_EMA_5
42.4661 ± 3.9131	Open_EMA_15
39.6315 ± 2.7256	Open_EMA_10
1.7036 ± 0.1859	Open_MACD
1.0491 ± 0.1013	Open_EMA_20
0.1780 ± 0.0076	Open_SMA_15
0.0483 ± 0.0041	Open_MACD_Signal
0.0262 ± 0.0009	Open_SMA_10
0.0185 ± 0.0011	Open_SMA_5
0.0031 ± 0.0002	Open_SMA_20
0.0000 ± 0.0000	Nasdaq_Open
0.0000 ± 0.0000	Year
0 ± 0.0000	Nasdaq_Open_Fluctuation_Rate
0 ± 0.0000	Nasdaq_Open_Volatility
0 ± 0.0000	Open_Fluctuation_Rate
0 ± 0.0000	Day
0 ± 0.0000	Month
0 ± 0.0000	Open_MACD_Oscillator
-0.0000 ± 0.0000	Open_Volatility



6. Conclusion

- Increasing Trend