JPX Tokyo Stock Exchange Prediction: Tokyo Gas Study Case

This project was built by a team of four persons including Azizha Zeinita

Background

- Japan Exchange Group, Inc. (JPX) is a holding company operating one of the largest stock exchanges in the world, which one of it’s significant stocks is Tokyo Gas stock.
- This project will compare several models against real future returns of Tokyo Gas stock after the training phase is complete.
- Other external datasets that are closely related to stock price, such as Japan GDP, Unemployment Rate, Oil Price, etc., will also involved in different combinations to find the best output.

Agenda

1. Business case and problem statement
2. Data description and properties
3. Data processing and EDA
4. Arima Models
5. Var Models
6. Prophet / LSTM Models
7. Insights/Recommendations & Future work

Business Case and Research Purpose

Select the **best performance model** that will be used for building the stock portfolio in the future and try to rank four energy stocks based on the future return.

Compared multivariate models with univariate models to determine which model produce higher accuracy score for stock price prediction.

Data Description and Properties

Stock Data:

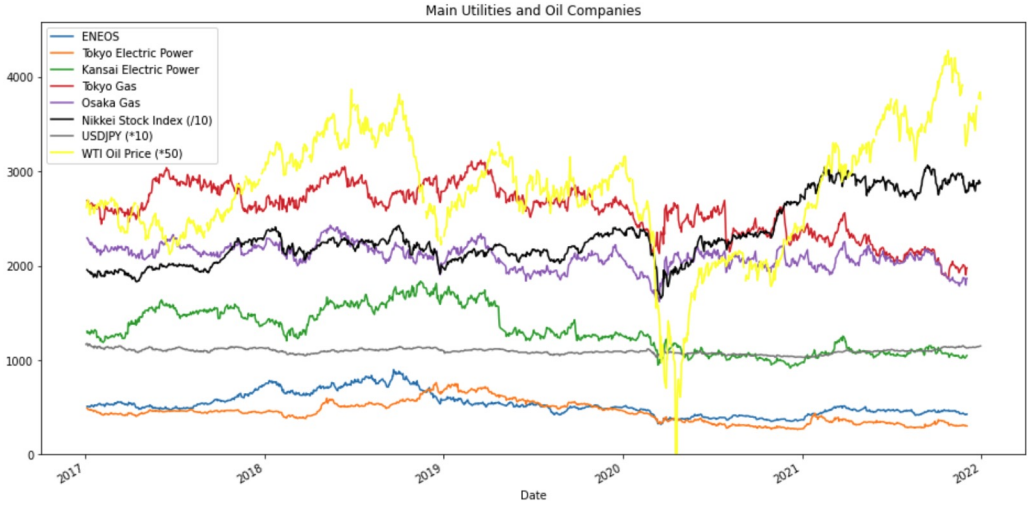
- Secondary Stock Price: Daily closing price for each stocks (**The Core**)
- Stock List: Company names and industry the company is in. (**EDA**)

Alternative Data (**Multivariate Models**):

- Japan Seasonally Adjusted Gross Domestic Product (GDP)
- Japan Seasonally Adjusted Unemployment Rate
- Daily Crude Oil Price
- Daily Exchange Rate for USD JPY
- Japan Daily 10 year Treasury Bond Interest Rate
- Nikkei 225 index

Exploratory Data Analysis

- Stocks are divided into 33 major sectors. Energy sector only has 38 companies.
- According to the **Random Walk Theory,** stock price are entirely random and there is no seasonality or trend.
- The trend of oil companies stock price **is similar to** the trend of the global oil price and the Nikkei 225 Index. In addition, oil companies stocks price **are correlated with** each other



Feature Engineering: External Regressors

Prepare 6 external regressors that are closely related to stock prices.

(1) Nikkei 225 Index(Daily) * Equivalent to S&P500

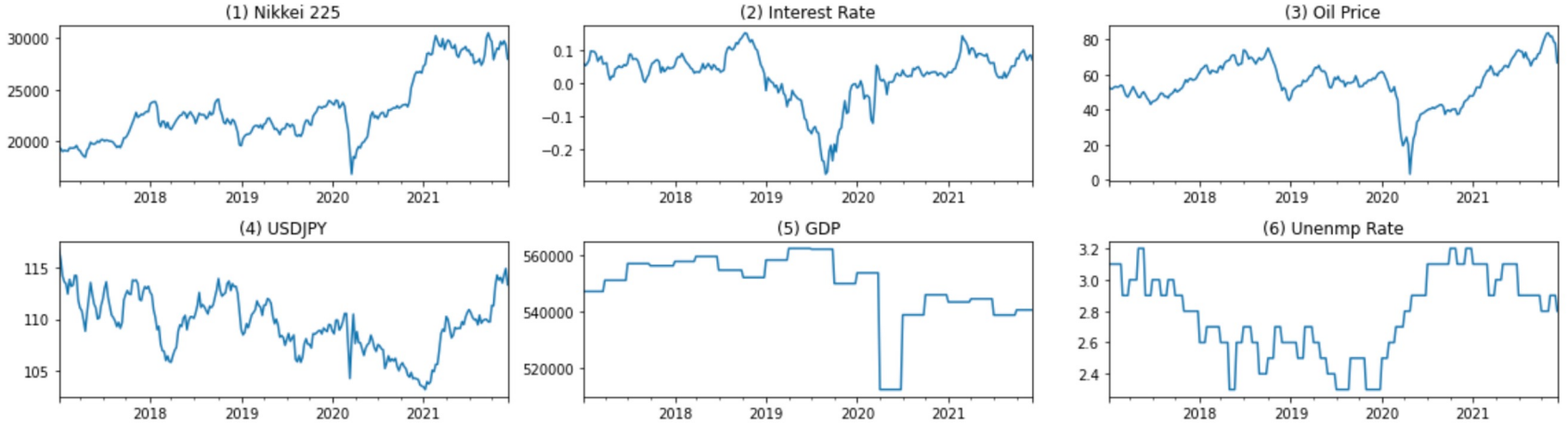
(2) Long-term Interest Rate(Daily) *For 10-year Treasury Bond

(3) Crude Oil Price(Daily) *West Texas Intermediate (WTI)

(4) Exchange Rate(Daily) *Yen/USD

(5) GDP(Quarterly) *Nominal GDP

(6) Unemployment Rate(Monthly) *For Aged 15-64

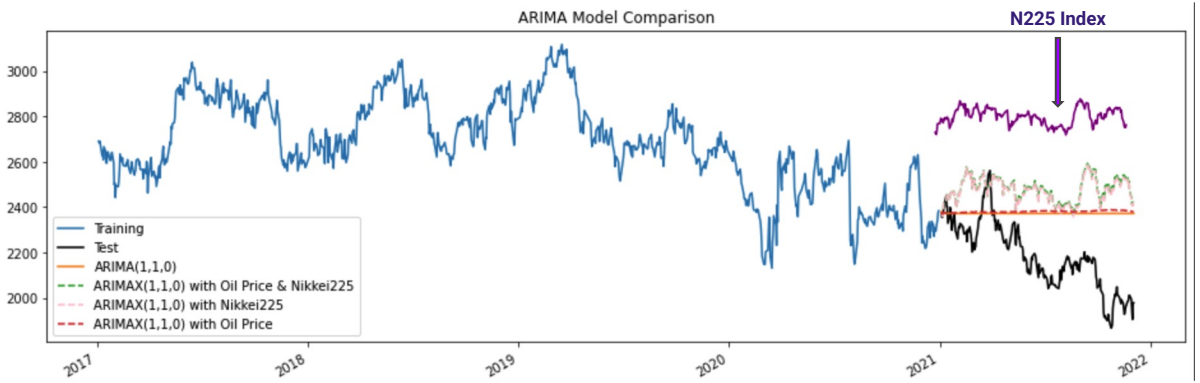


Models

ARIMA, VAR, PROPHET, LSTM

ARIMA Model Daily

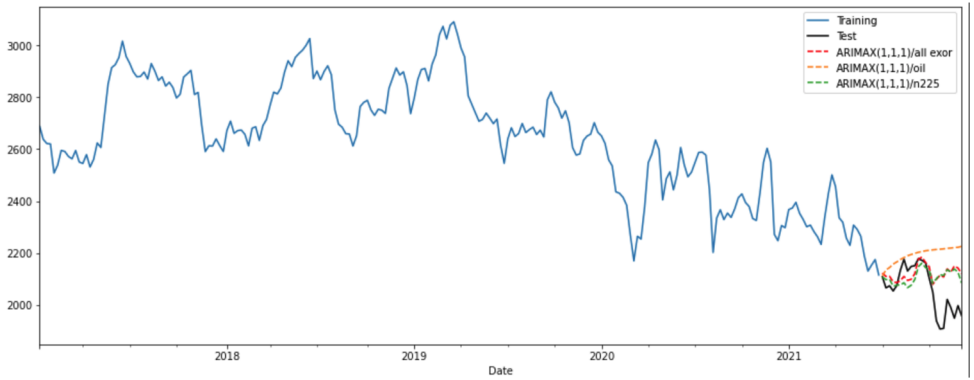
- All ARIMA models performed poorly on predicting the price
- Univariate ARIMA does generate meaningful outcome
- Multivariate ARIMA captured too much trend of N225 index



	MAE	MSE	MAPE	sMAPE	MASE
Univariate	203.792	58725.1	0.0978	0.0914	-8.7564
Oil	211.5135	63002.5703	0.1015	0.0946	-9.0882
N225	292.0734	107173.8052	0.1389	0.1274	-12.5497
Oil, N225	298.0109	111508.5736	0.1417	0.1298	-12.8048

ARIMA Model Weekly

- Shorten the prediction time and using weekly mean helps to improve ARIMA accuracy
- Univariate ARIMA does generate meaningful outcome
- Multivariate ARIMA captured too much trend of N225 index

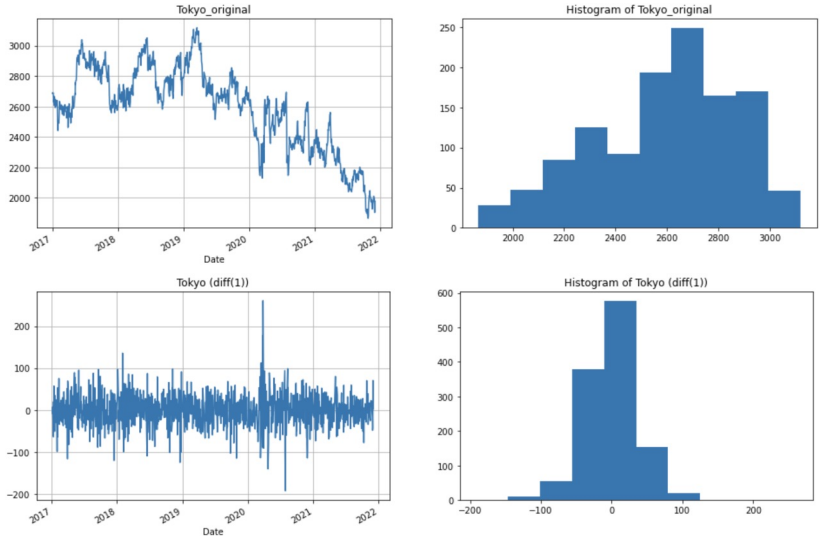


The worst
The best

	MAE	MSE	MAPE	sMAPE	MASE
Univariate	131.9129	28228.7658	0.0661	0.0627	-1.1125
Oil	129.9804	27404.4503	0.0652	0.0618	-1.0962
N225	78.5779	10447.2298	0.0393	0.0382	-0.6627
All	77.2154	10671.9102	0.0388	0.0375	-0.6512

VAR Model

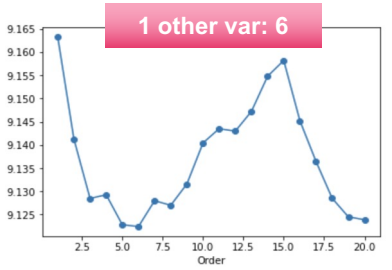
- Many real-world forecasting problems have **more than one variable** to consider.
- Vector ARIMA models can help model **multivariate** time series where:
  - The variables are **serially correlated**
  - The variables are **cross-correlated**
- VAR in **levels** – using stationary data
- VAR in **differences** – using differenced data
- We use 6 different external variables:
  - **Oil, N225, Unemploy, GDP, USD-JPY, interest**

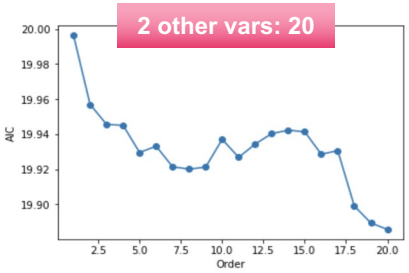


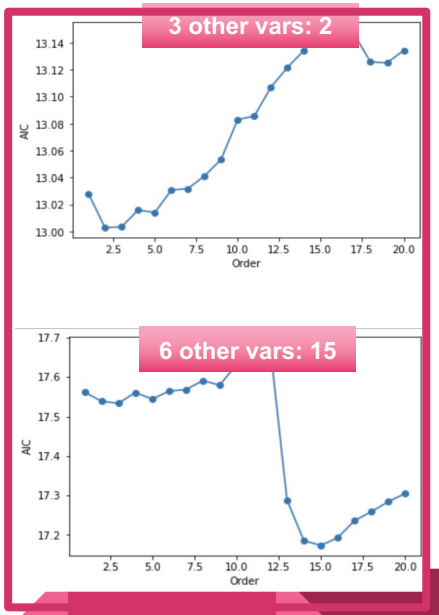
VAR Model

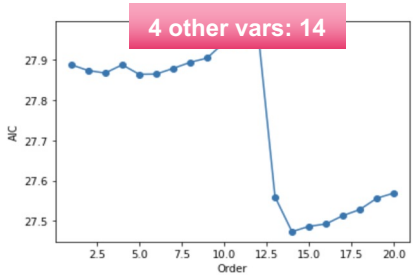
AIC Graphs Comparison

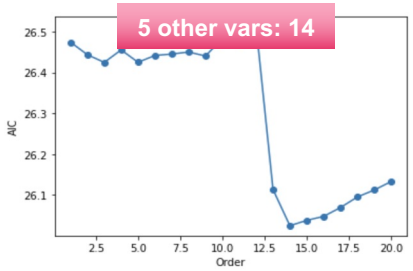
- Range of the lowest AIC: **2 - 20**
- The **best** model: VAR(15)
- The **worst** model: VAR (2)







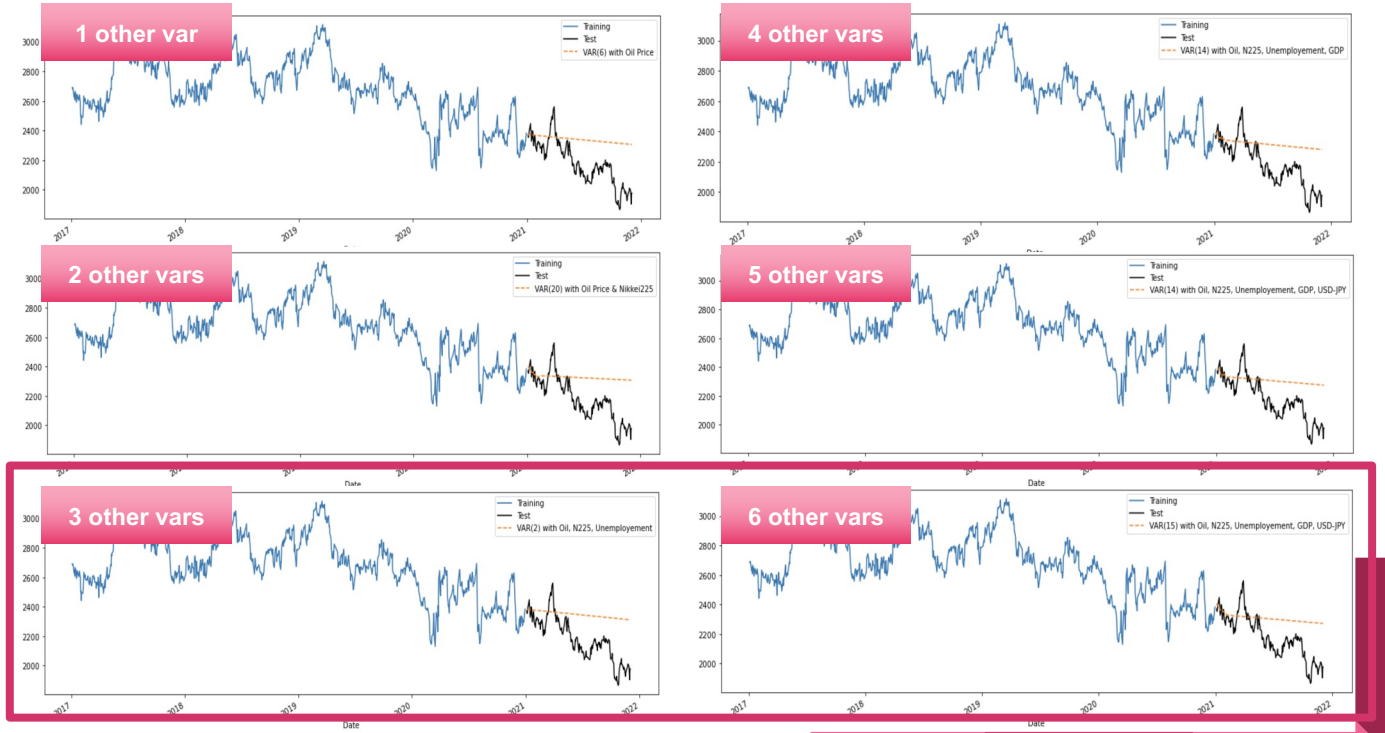




VAR Model

Forecasting Model Graphs Comparison

- All predictions are visually **almost similar**
- Even the worst (3 ex. var) and the best (6 ex. var) are **slightly different**
- Closer at the beginning, **further away in the ⅔** of the pred. period.



VAR Model

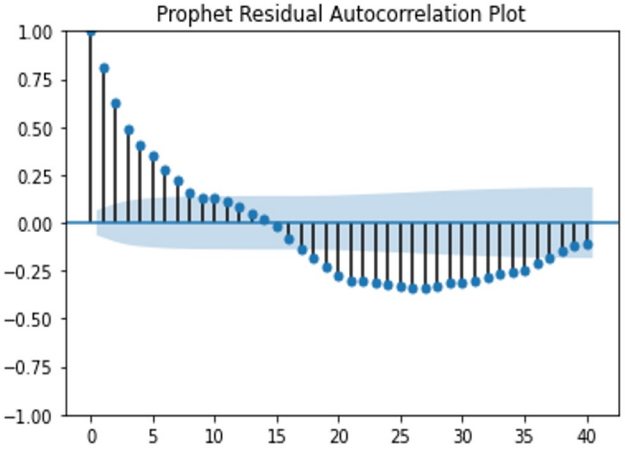
Test Error Metrics

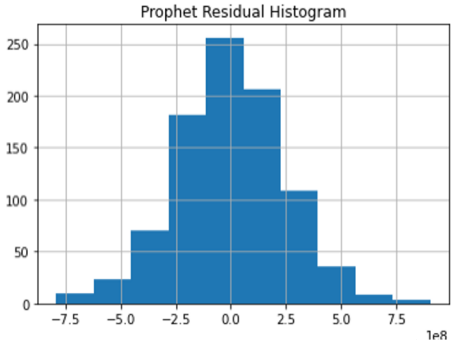
The worst
The best

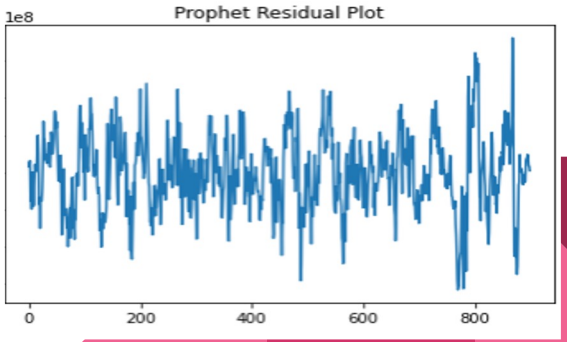
	MAE	MSE	MAPE	sMAPE	MASE
Oil	173.6549	42873.2797	0.0833	0.0786	-7.4615
Oil, N225	164.9285	40503.7034	0.0793	0.0749	-7.0866
Oil, N225, Unemploy	179.2476	44981.3307	0.0859	0.081	-7.7018
Oil, N225, Unemploy, GDP	155.0108	35544.422	0.0744	0.0706	-6.6604
Oil, N225, Unemploy, GDP, USD-JPY	149.9029	33680.3447	0.072	0.0684	-6.441
Oil, N225, Unemploy, GDP, USD-JPY, interest	148.0882	32951.911	0.0711	0.0676	-6.363

Prophet Model

- Univariate Analysis
- An additive model where linear or non-linear trends are fit to the data with seasonality.
- Model Fitting
  - Trained the model by using hyperparameter-tuning
    - changepoint_prior_scale=0.5
    - seasonality_prior_scale=0.01
- Residual Checks:
  - Prophet TRAIN MAE: 48.20
  - Prophet TRAIN MSE: 4135.62
  - Prophet TRAIN MAPE: 0.02
  - Prophet TRAIN SMAPE: 0.02

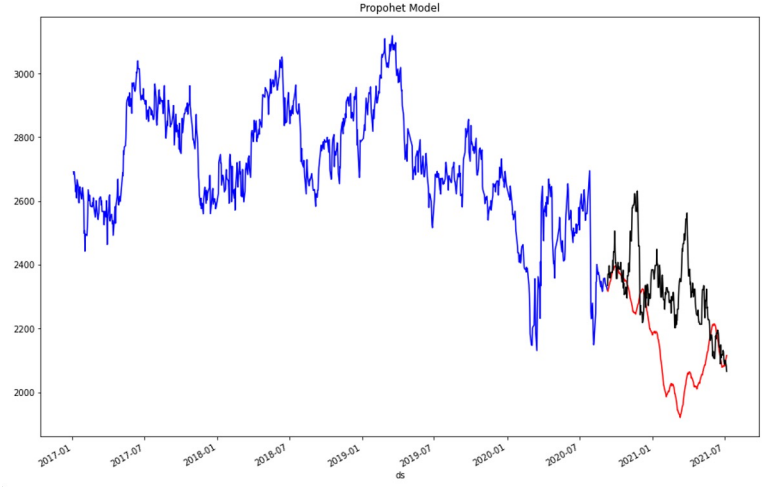


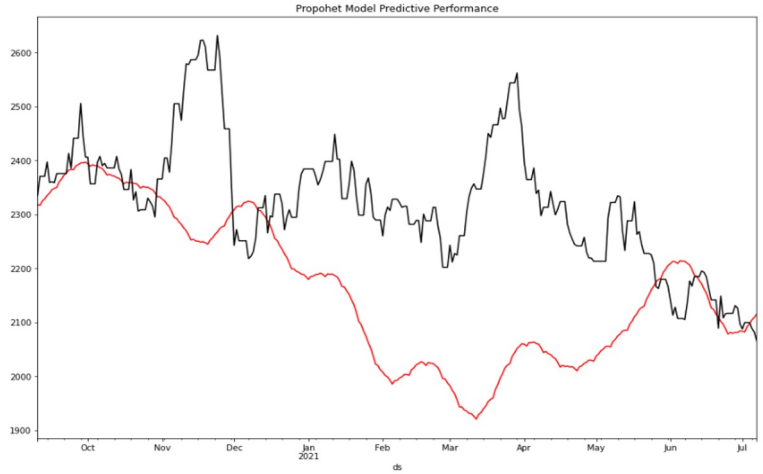




Prophet Model

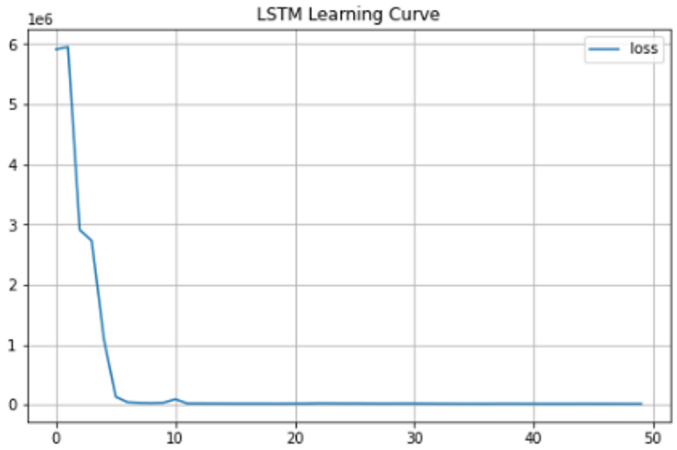
- Predictive Accuracy (Forecasting Period: 09/10/2020 - 12/03/2021):
  - Prophet TEST MAE: 174.70
  - Prophet TEST MSE: 49763.55
  - Prophet TEST MAPE: 0.07
  - Prophet TEST SMAPE: 0.08

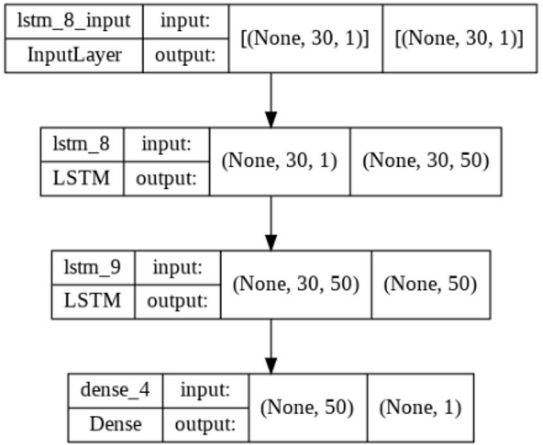




Long Short-Term Memory (LSTM)

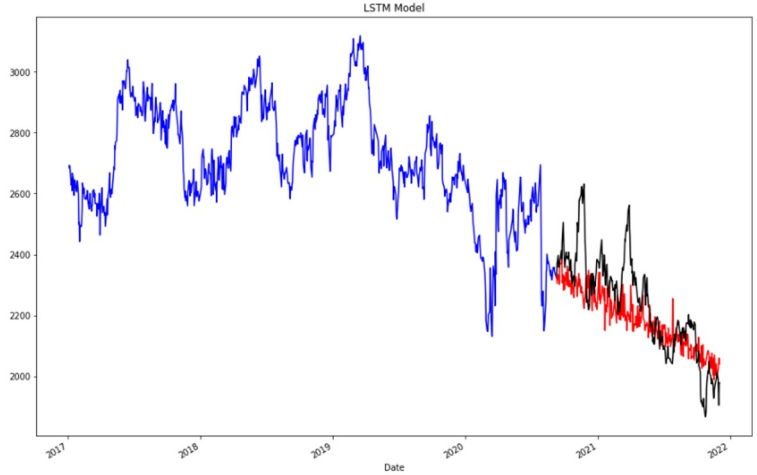
- Univariate Analysis
- The network can learn what to store in the long-term state, what to throw away, and what to read from it.
- Model Fitting:

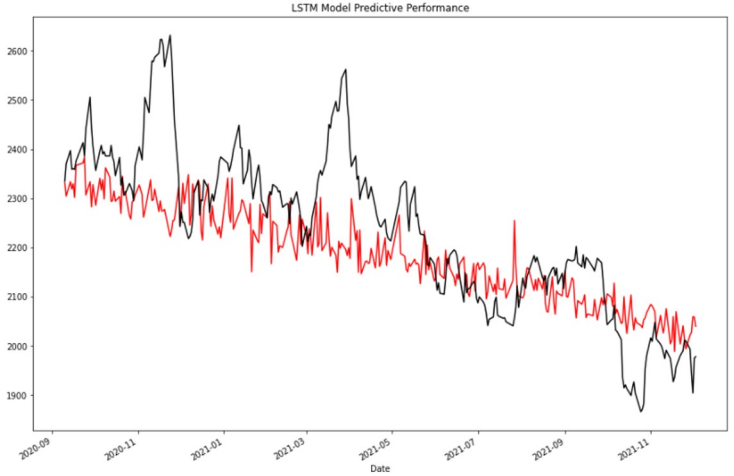




Long Short-Term Memory (LSTM)

- Predictive Accuracy (Forecasting Period: 09/10/2020 - 12/03/2021):
  - LSTM TEST MAE: 88.87
  - LSTM TEST MSE: 14273.59
  - LSTM TEST MAPE: 0.04
  - LSTM TEST SMAPE: 0.04





Conclusion
&
Insights

Model Comparison

The worst
The best

	MAE	MSE	MAPE	SMAPE
ARIMA (Univariate)	203.792	58725.1	0.10	0.09
VARIMA (Oil, N225, Unemploy, GDP, USD-JPY, interest)	148.09	32951.91	0.07	0.07
Prophet	174.70	49763.55	0.07	0.08
LSTM	**88.87**	**14273.59**	**0.04**	**0.04**

Insights and Future Work

1. Predictive accuracy is ***high in the short-term***. In the ***long-term***, the predictions are ***less accurate***.
2. Build models for all 2000 stocks and analyze the performance.
3. Include more features in these models.
4. Apply more Machine Learning Algorithms: Regression model, Decision Tree, CNN