Capstone project

STI price forecasting using time series machine learning models

Prepared by: Miko Poh Chin Hong

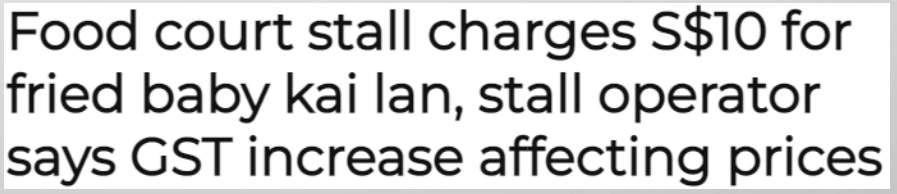
Problem statement

Financial institutes offering STI can only project returns of around 5% per annum whereas Singapore current inflation rate is more than 6%

- Market has been volatile in recent years due to pandemic, war and interest rate hike
- Aim to have higher returns than inflation rate using machine learning models
- Singapore index which is called Straits Times Index (STI) consist of 30 components/stocks
- Market index is a measurement of subset of stock market as a value
- Why index being chosen for analysis:
  - Broad diversification
  - Lowered risk
  - Able to track 24/5

From https://dollarsandsense.sg/singapores-inflation-remained-steady-6-7-november/

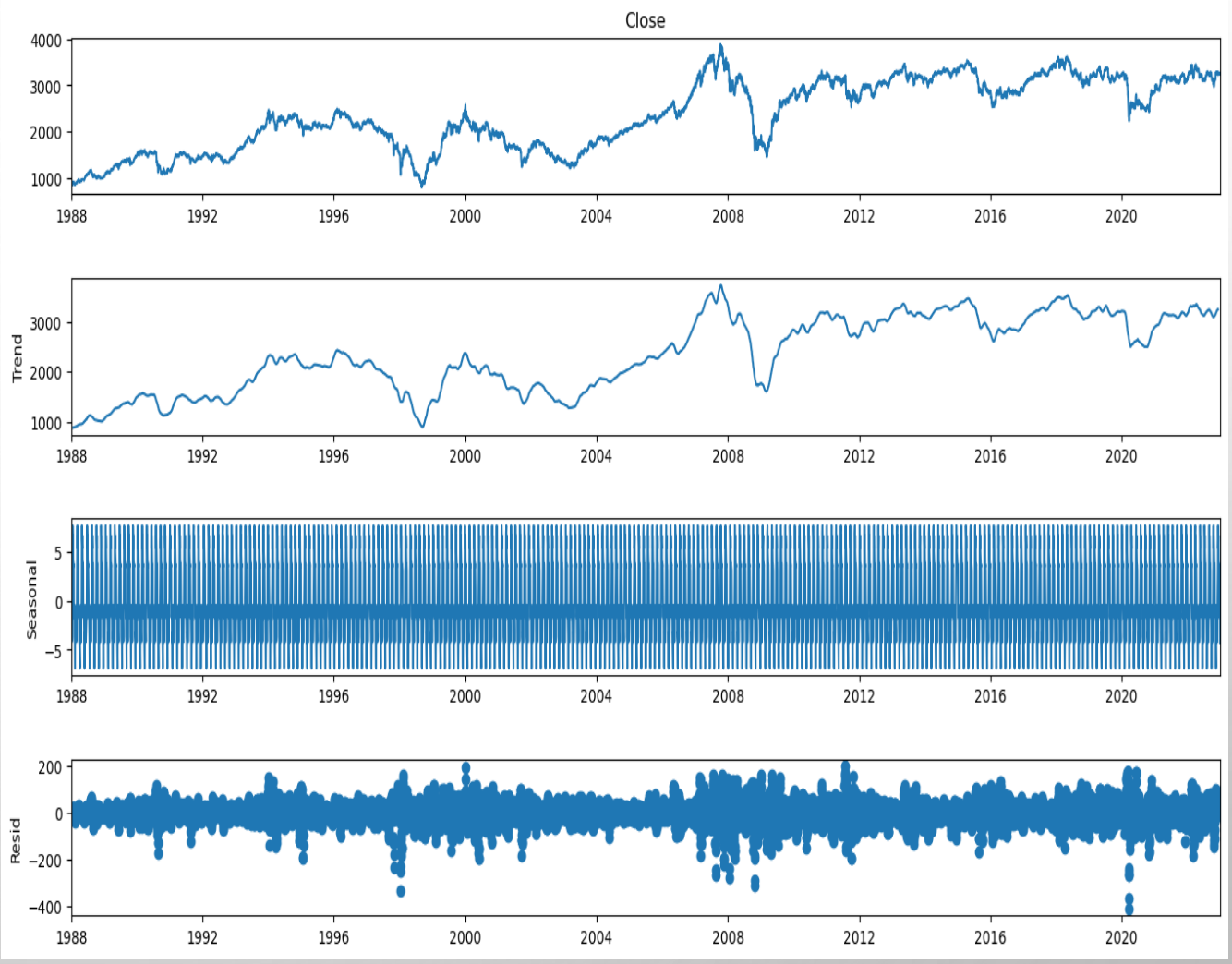
From https://www.channelnewsasia.com/singapore/singpost-adjust-increase-rates-prices-postage-gst-2023-3149501



From https://theindependent.sg/food-court-stall-charges-s10-for-fried-baby-kai-lan-stall-operator-says-gst-increase-affecting-prices/

Time series analysis and forecasting

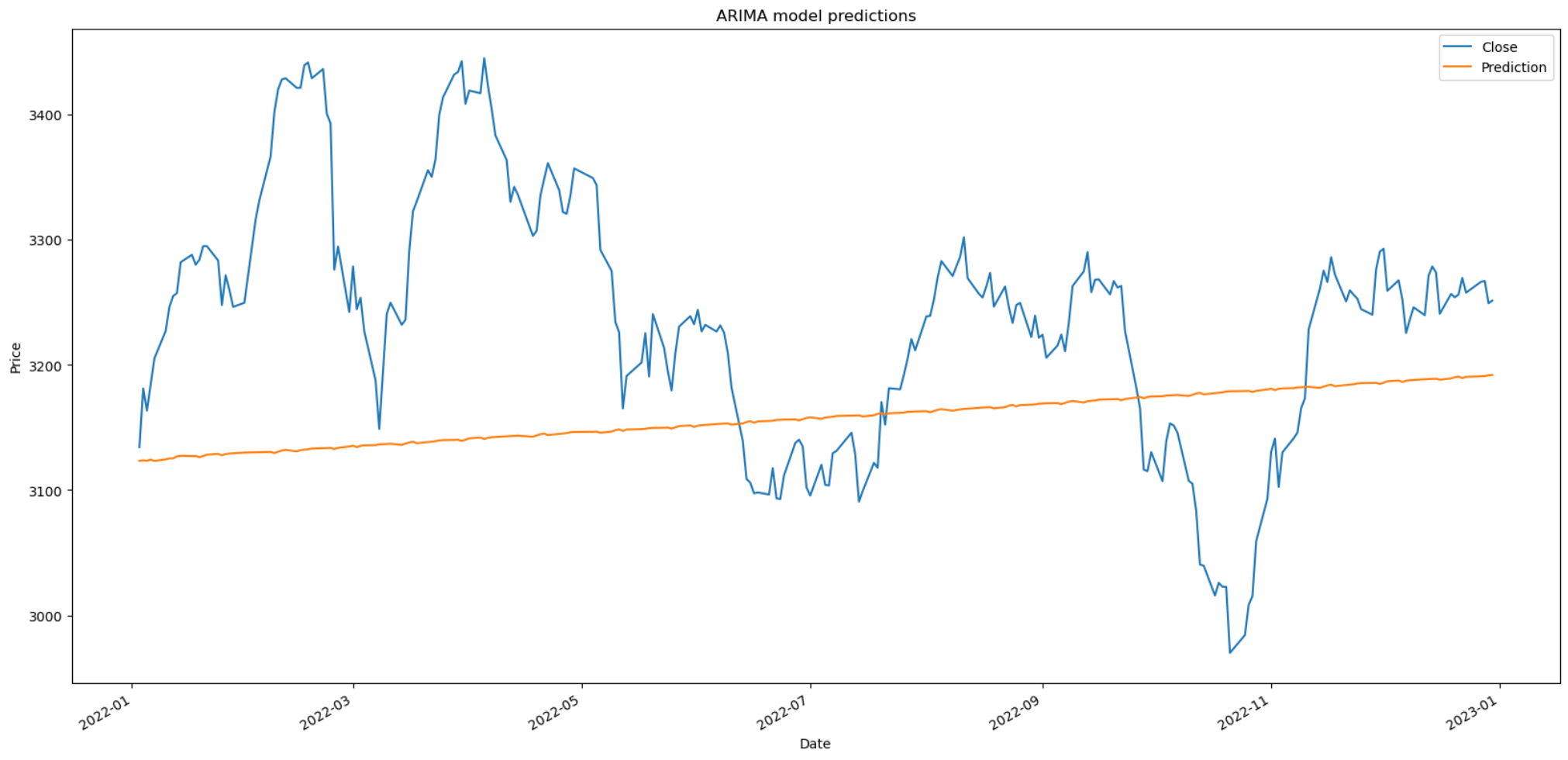
- Time series data are recorded at regular intervals over a specified period of time
- Time series analysis used for non-stationary data
  - Stock prices
  - Weather
  - Temperature
- Two time series models used to forecast:
  - Autoregressive integrate moving average (ARIMA)
  - Long short term memory (LSTM)
- Autoregressive integrated moving average
  - AR – Autoregression refer to changing value that regresses on its lagged values
  - I – Integrated means to observed difference between static data values and pervious data values (stationary data)
  - MA – Moving average represent dependency between observed value and residual error



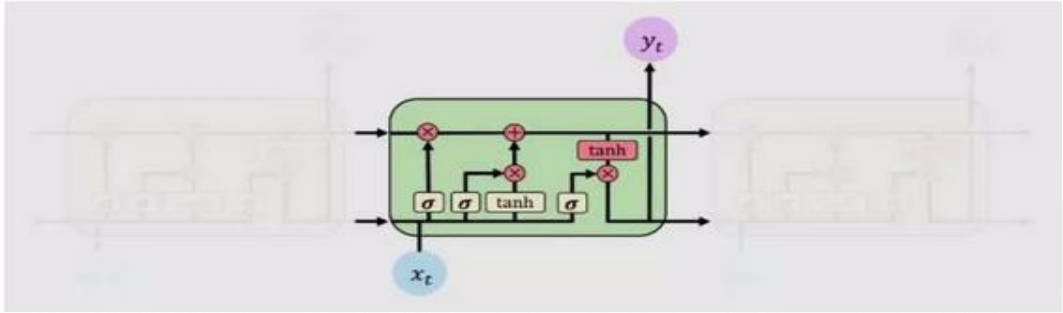
Error-Trend-Seasonality (ETS) model for STI closing price

SARIMAX prediction of 2022 STI closing price

- SARIMAX is able to predict overall trend
- Unable to predict well for long periods
- Not suitable for this set of time series or duration
- Proceed to use long short term memory (LSTM) for prediction



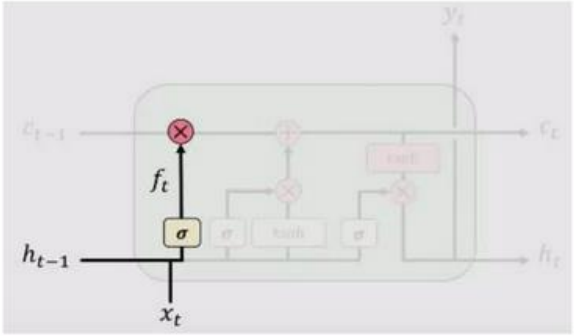
Long short term memory (LSTM)



- A subtype of recurrent neural network(RNN) layered structure of connected neurons
- Typical recurrent Neural Networks have a short-term memory to retain previous information in the current neuron
- LSTM on the other hand have long and short term memory
- LSTM is also able to use features (multivariate)

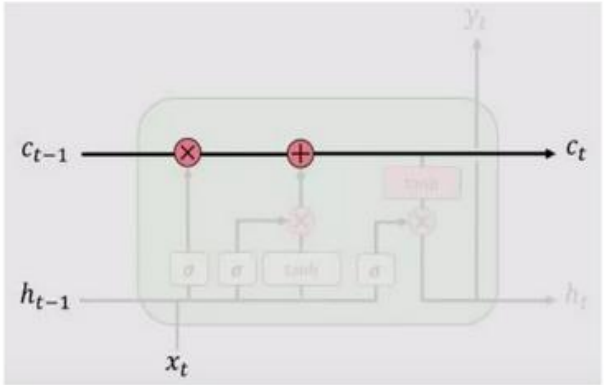
1) Forget 2) Store 3) Update 4) Output

LSTMs forget irrelevant parts of the previous state



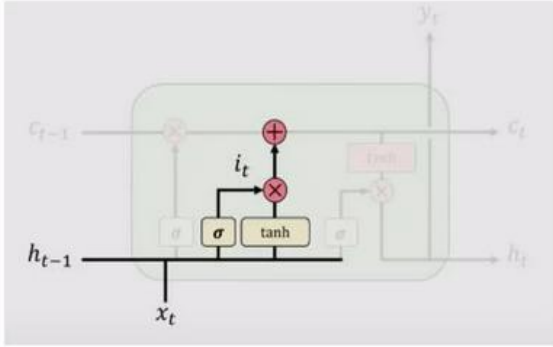
1) Forget 2) Store **3) Update** 4) Output

LSTMs selectively update cell state values



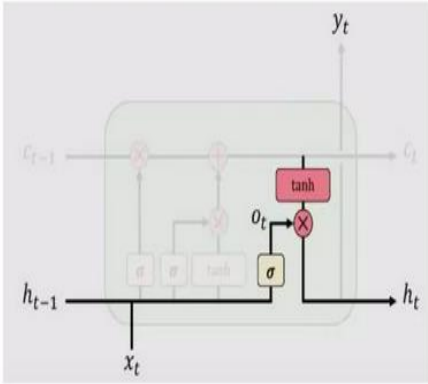
1) Forget **2) Store** 3) Update 4) Output

LSTMs store relevant new information into the cell state



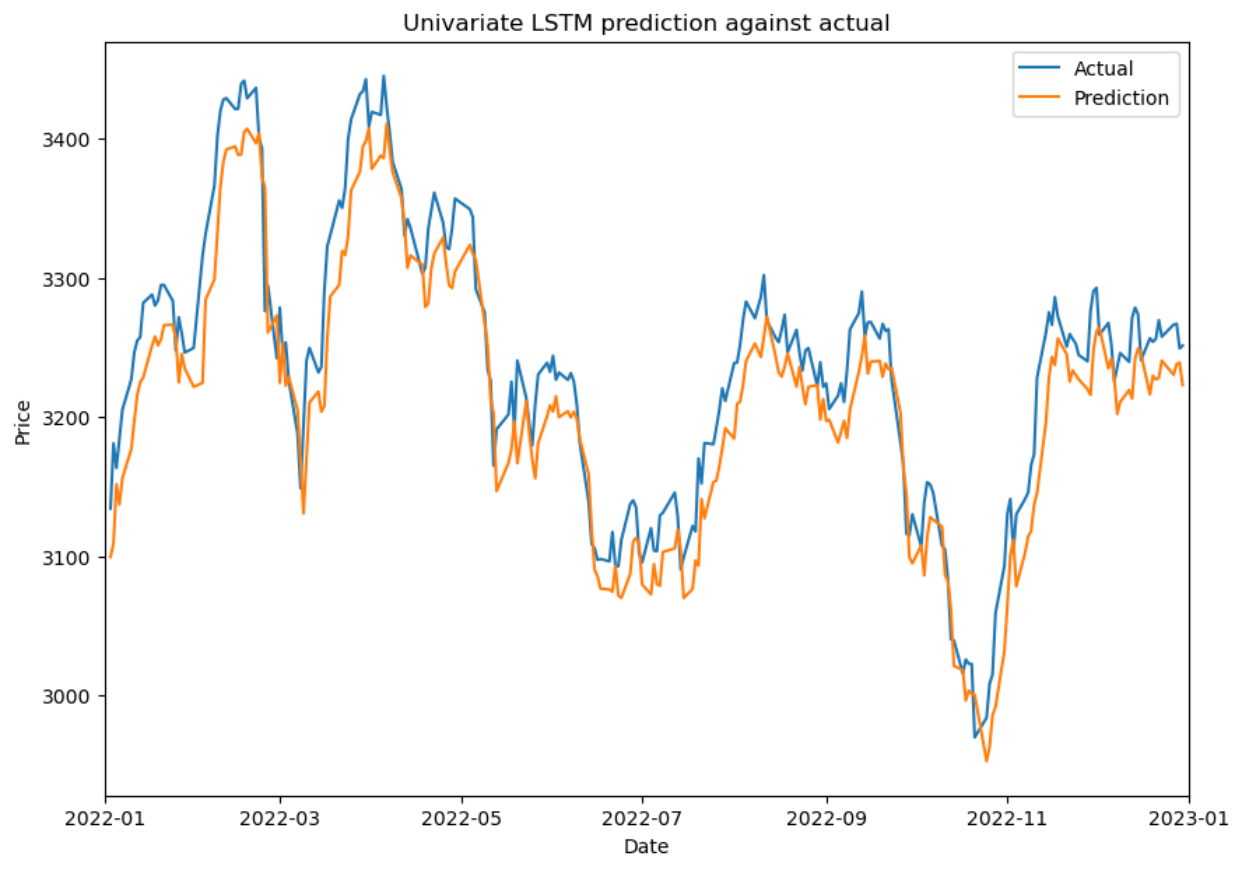
1) Forget 2) Store 3) Update **4) Output**

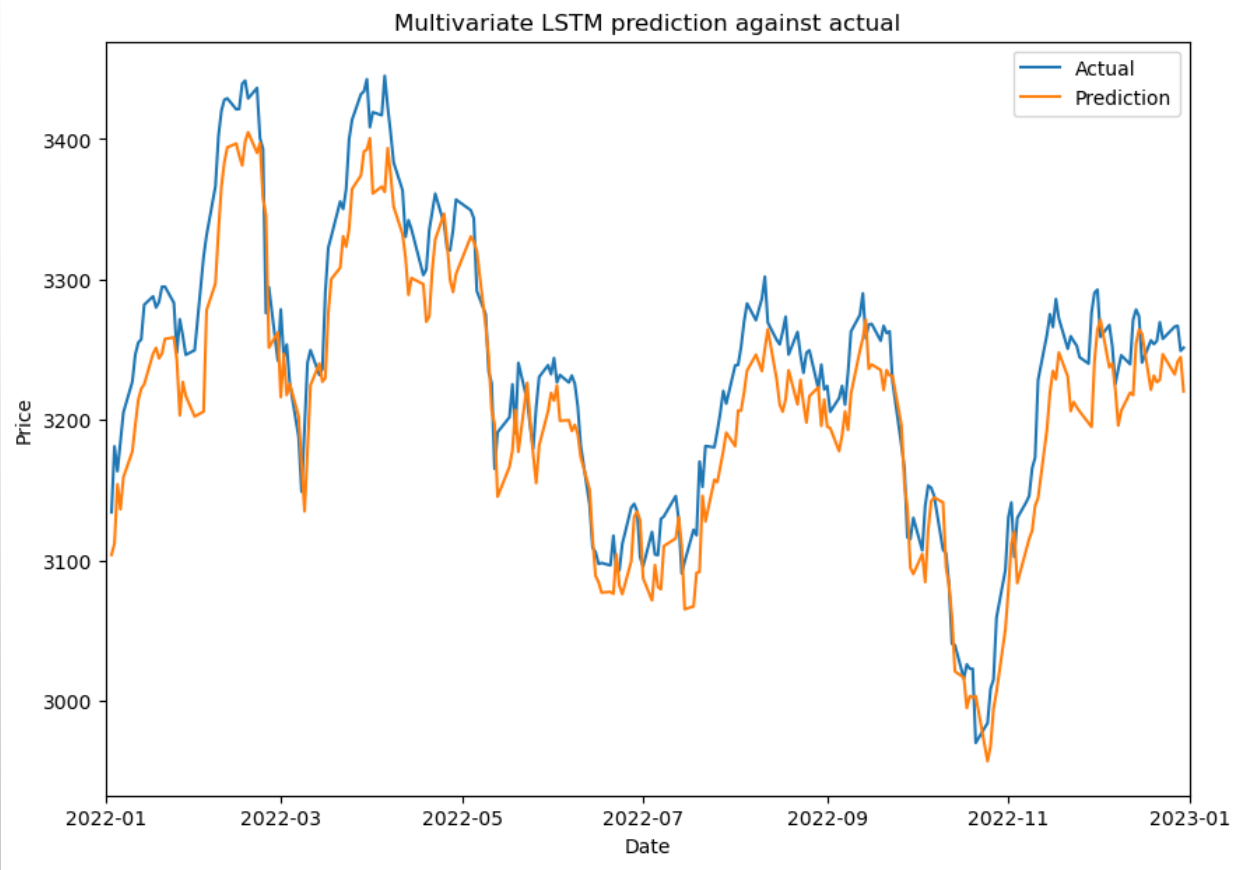
The output gate controls what information is sent to the next time step



LSTM predictions against actual of STI closing prices for 2022

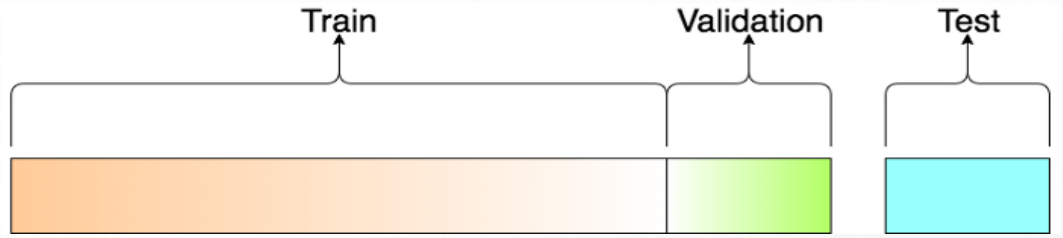
- Two different LSTM models are tested: Univariate and multivariate
- Both univariate and multivariate LSTM's prediction are able to follow closely to actual closing price
- Not much difference at first glance
- In order to compare both models, loss curve and final MSE score is used



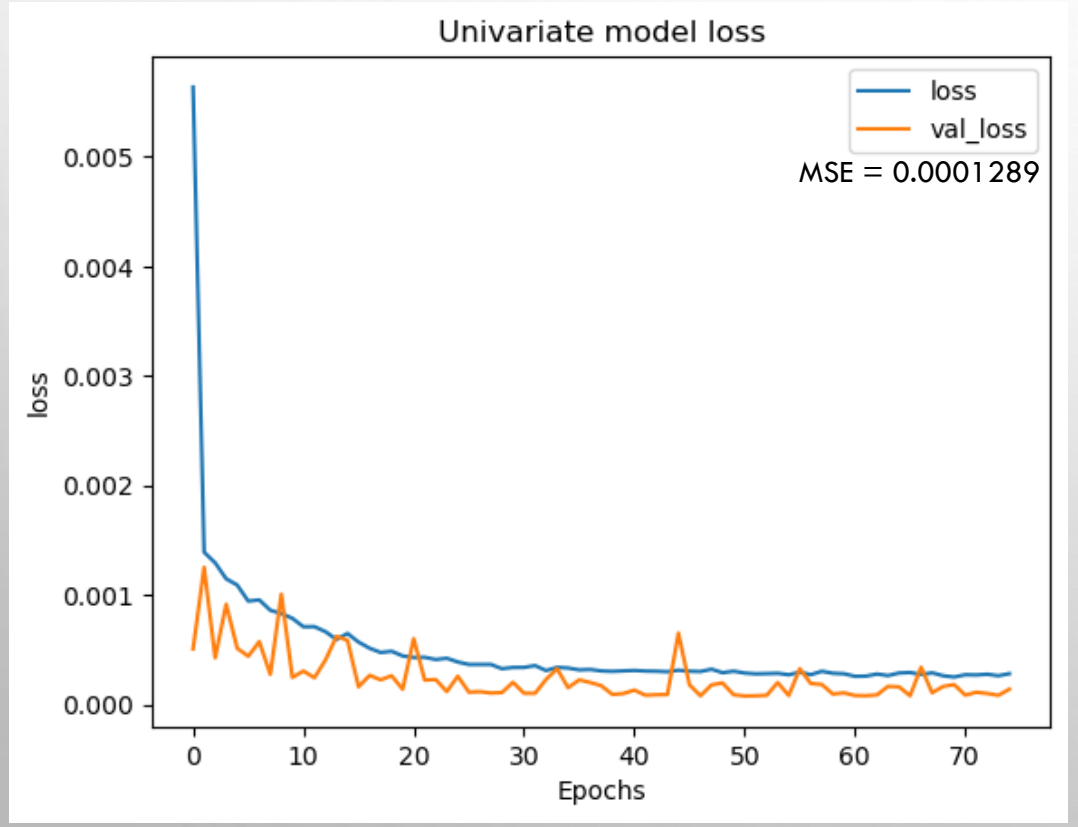


Loss curves of univariate and multivariate LSTM

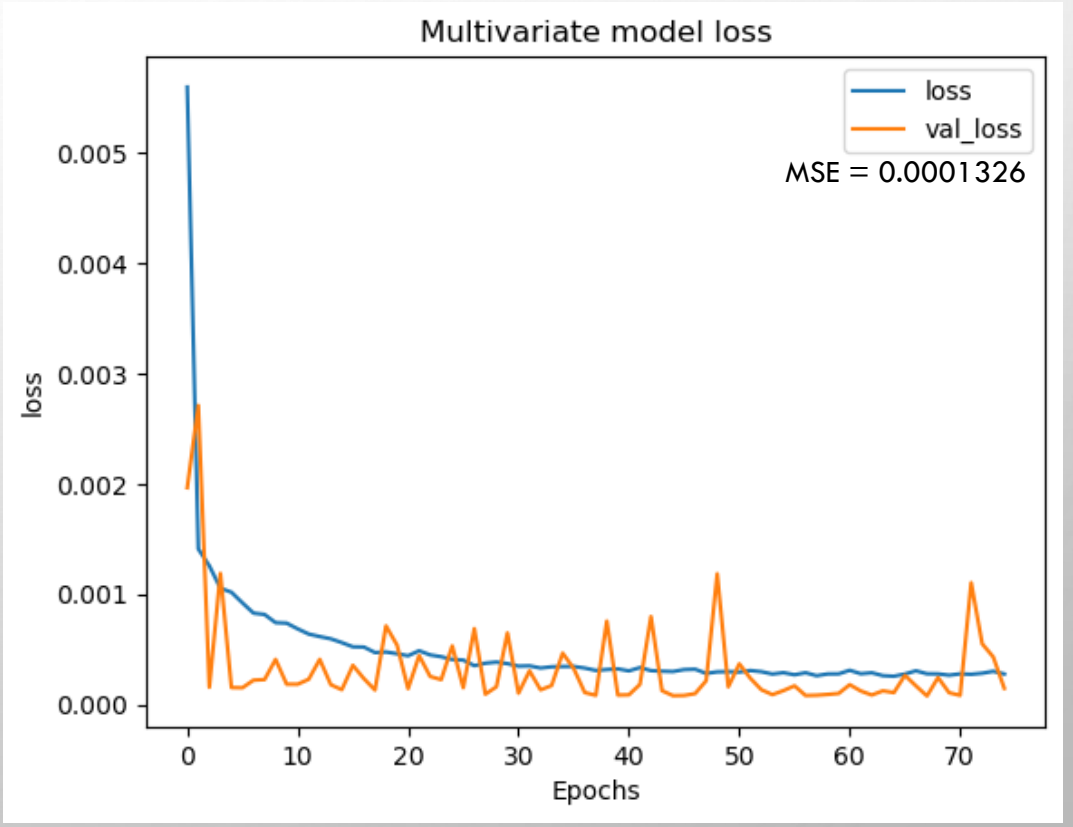
- A portion of train dataset is split as validation data set to generate a loss number, the validation set is not seen by the model during training
- For every epoch, a loss number is generated
- Base on MSE score, univariate LSTM model is better performing

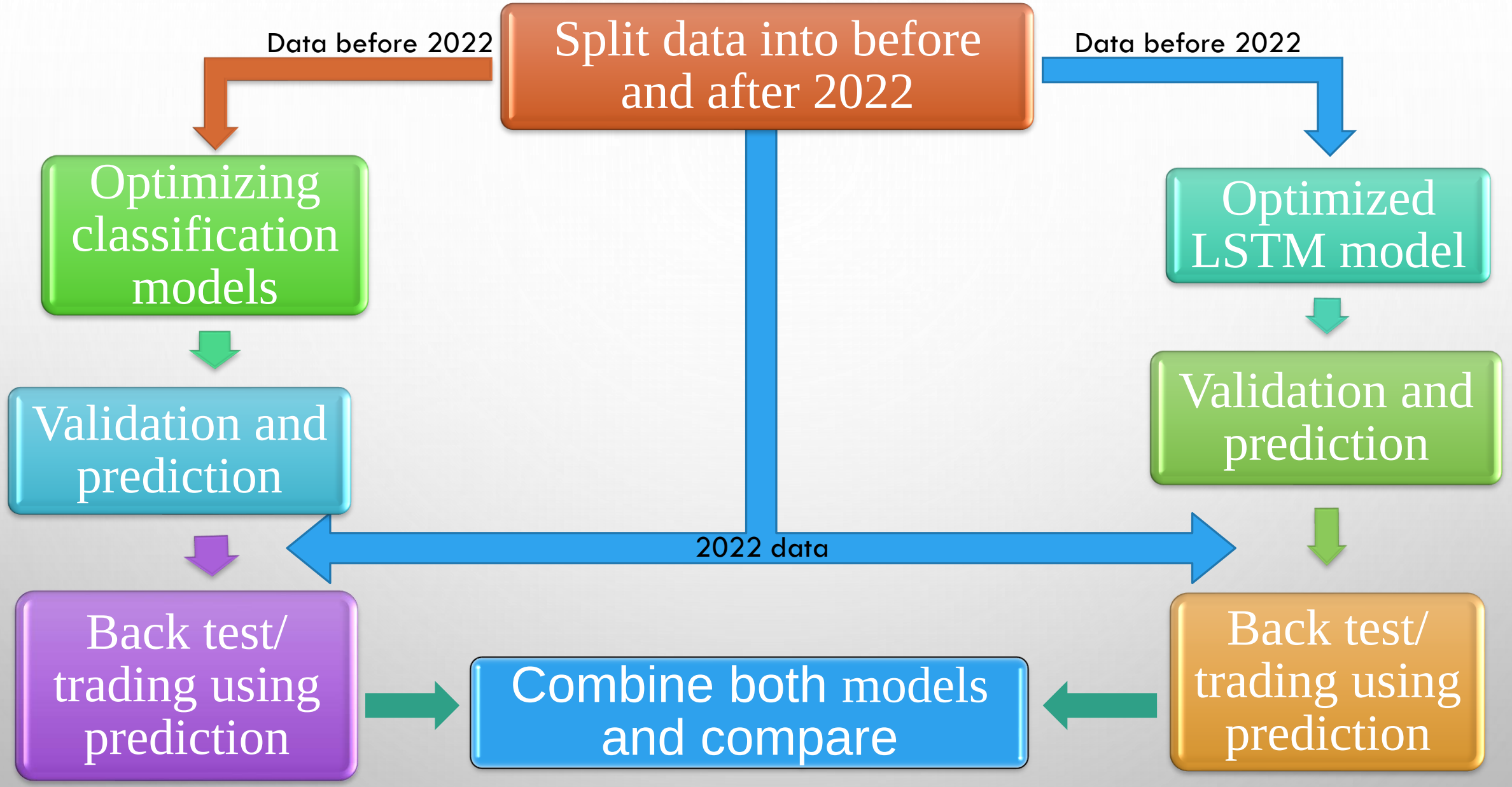


Loss of univariate LSTM



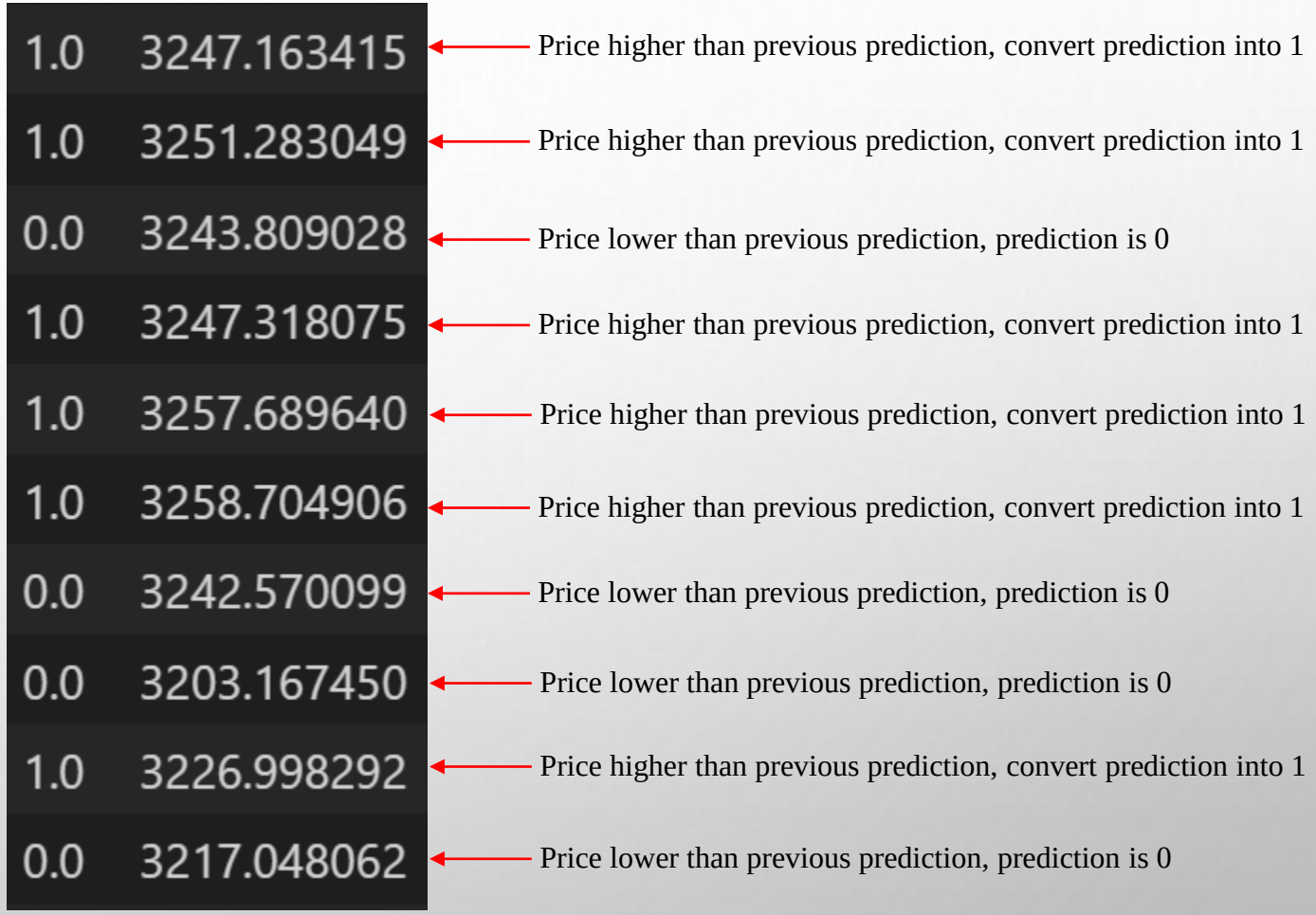
Loss of multivariate LSTM




Split data into before and after 2022
Data before 2022
Data before 2022
Optimizing classification models
Optimized LSTM model
Validation and prediction
Validation and prediction
2022 data
Back test/ trading using prediction
Combine both models and compare
Back test/ trading using prediction

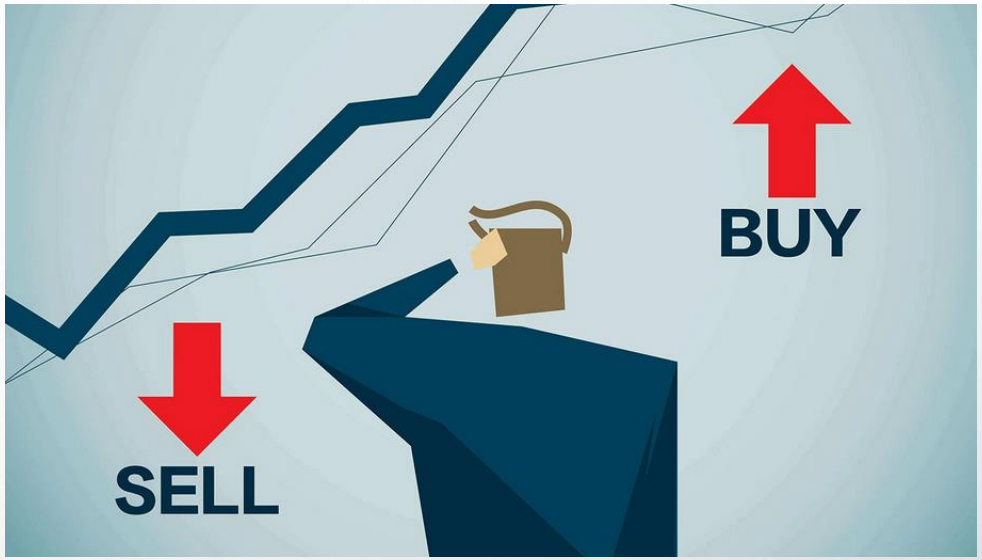
Combining LSTM predictions with classification machine learning models

- Predictions from ML classification
  - ➢ Technical indicators used as features:
    - ❖ Change in 5 days exponential moving average
    - ❖ Stochastic of 5 days
    - ❖ Stochastic of 10 days
    - ❖ Change in MACD
    - ❖ Force index of 5 days
  - ➢ ML model used is XGBoost model
  - ➢ Output is 1 for predicting up and 0 for predicting down
- Converting LSTM predictions to classification
  - ➢ If current day prediction is higher than previous day prediction, prediction is 1
  - ➢ If current day prediction is lower than previous day prediction, prediction is 0
- Combining of two or more models when all models give the same prediction



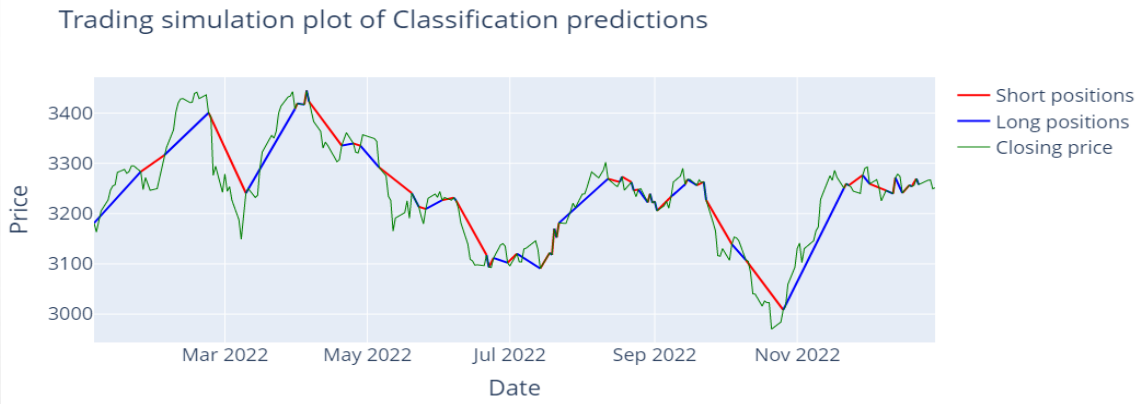
Back test/ trading simulation of 2022 data predictions

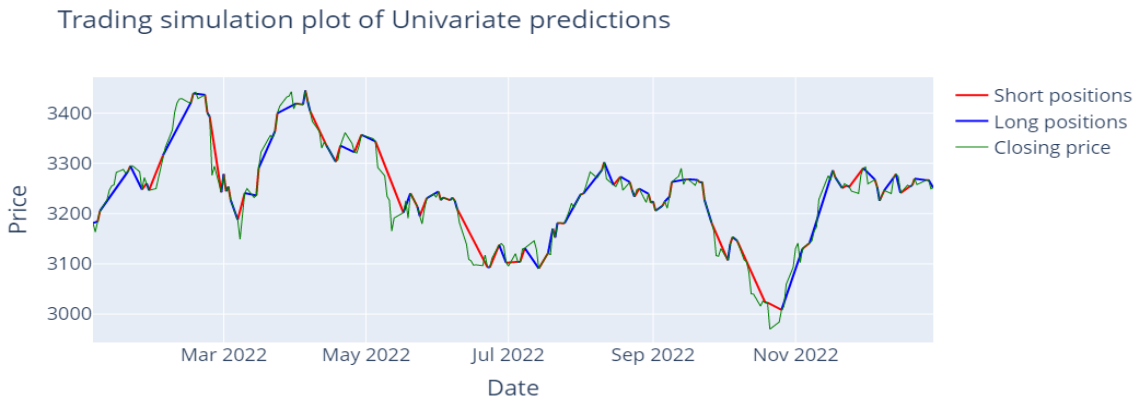
- Prediction is use to simulate trading scenarios with the following strategies and assumptions
  - Start of with zero position in market and first prediction 0 to sell and 1 to buy
  - If prediction and current position are the same or no prediction, hold position
  - If prediction is different from current position, current position is closed and enter new position is taken with the prediction direction
  - Assuming buying and selling price are exactly at closing price when the prediction appear

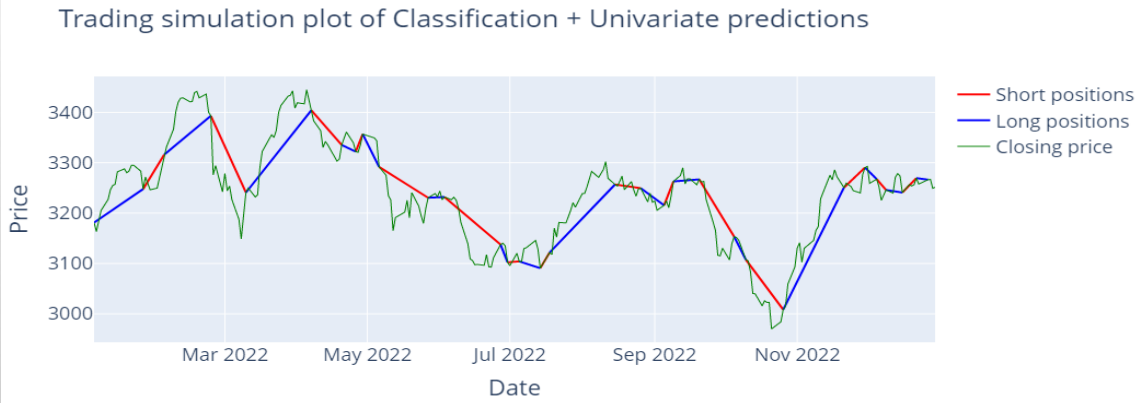


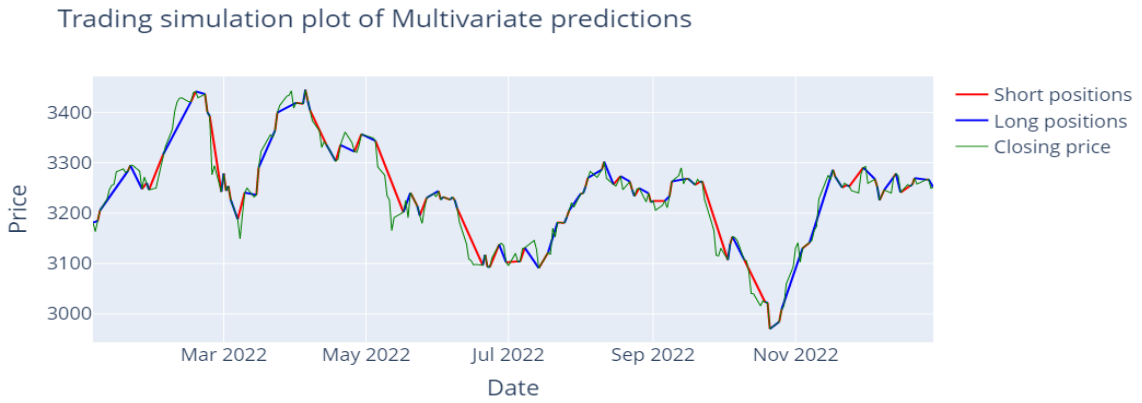
Date	Close	Next_day_movement	lstm_prediction	clf_prediction	combine_prediction	
2022-01-18	3280.040039	1.0	1.0	1	1.0	Up prediction , since starting with no position so buy at 3280.04
2022-01-19	3283.939941	1.0	0.0	1	no_prediction	No prediction, hold position
2022-01-20	3294.820068	1.0	1.0	1	1.0	Up prediction, hold position since already in a long position
2022-01-21	3294.860107	0.0	1.0	1	1.0	Up prediction, hold position since already in a long position
2022-01-24	3283.350098	0.0	1.0	0	no_prediction	No prediction, hold position
2022-01-25	3247.760010	0.0	0.0	0	0.0	Down prediction, close previous position and enter new short position
2022-01-26	3271.570068	0.0	0.0	0	0.0	Down prediction, hold position since already in a short position
2022-01-27	3260.030029	0.0	1.0	0	no_prediction	No prediction, hold position
2022-01-28	3246.330078	0.0	0.0	0	0.0	Down prediction, hold position since already in a short position
2022-01-31	3249.590088	1.0	0.0	0	0.0	Down prediction, hold position since already in a short position

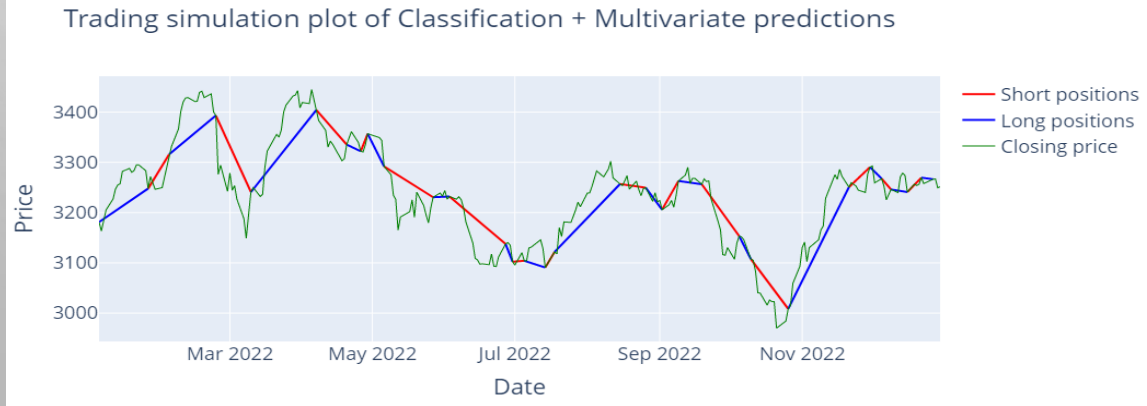
Trading actions taken based on model predictions

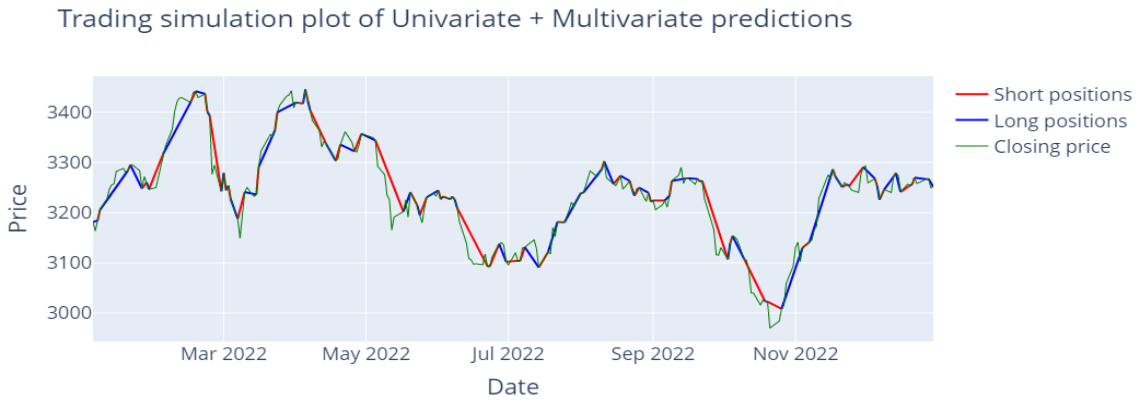




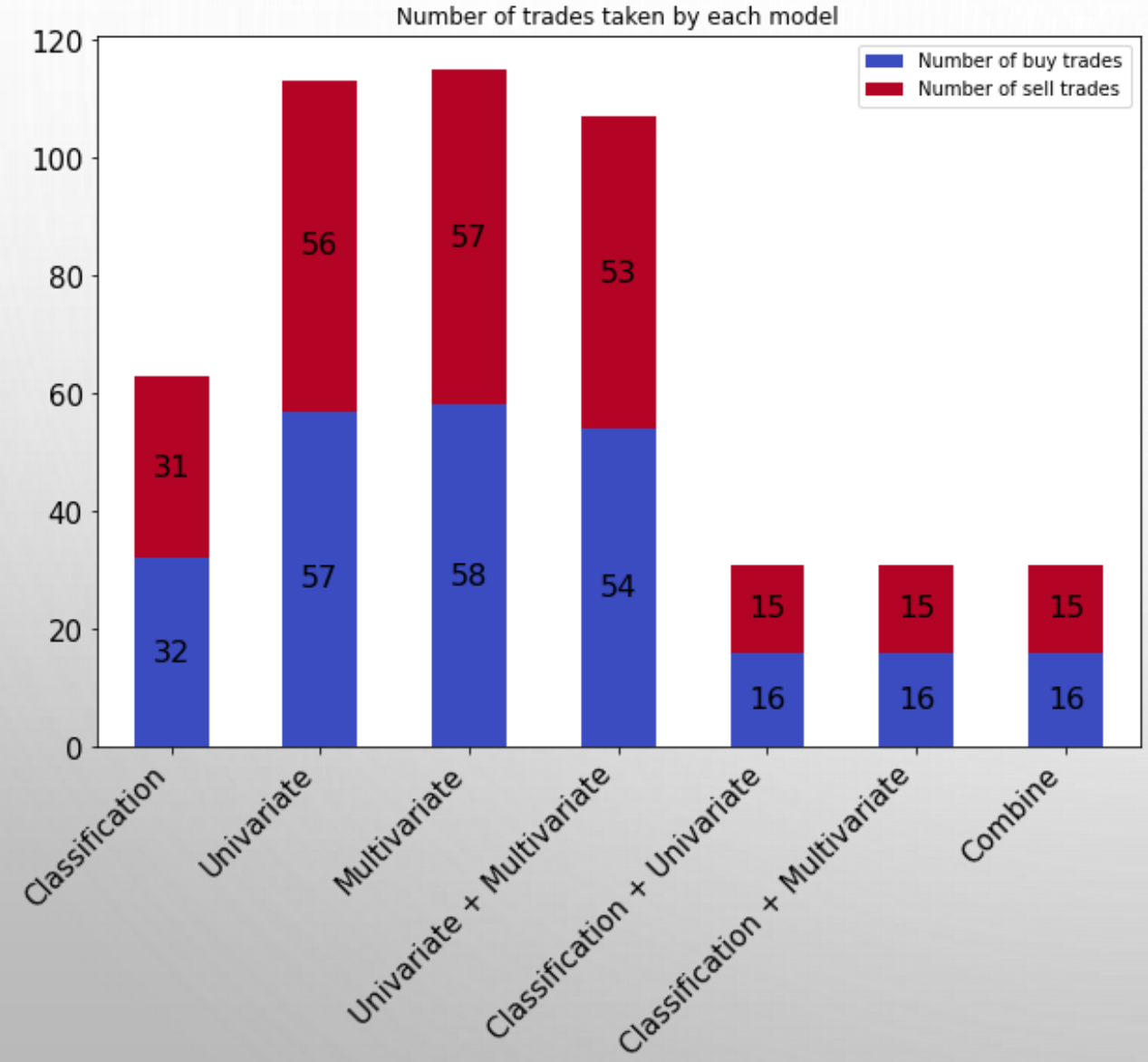


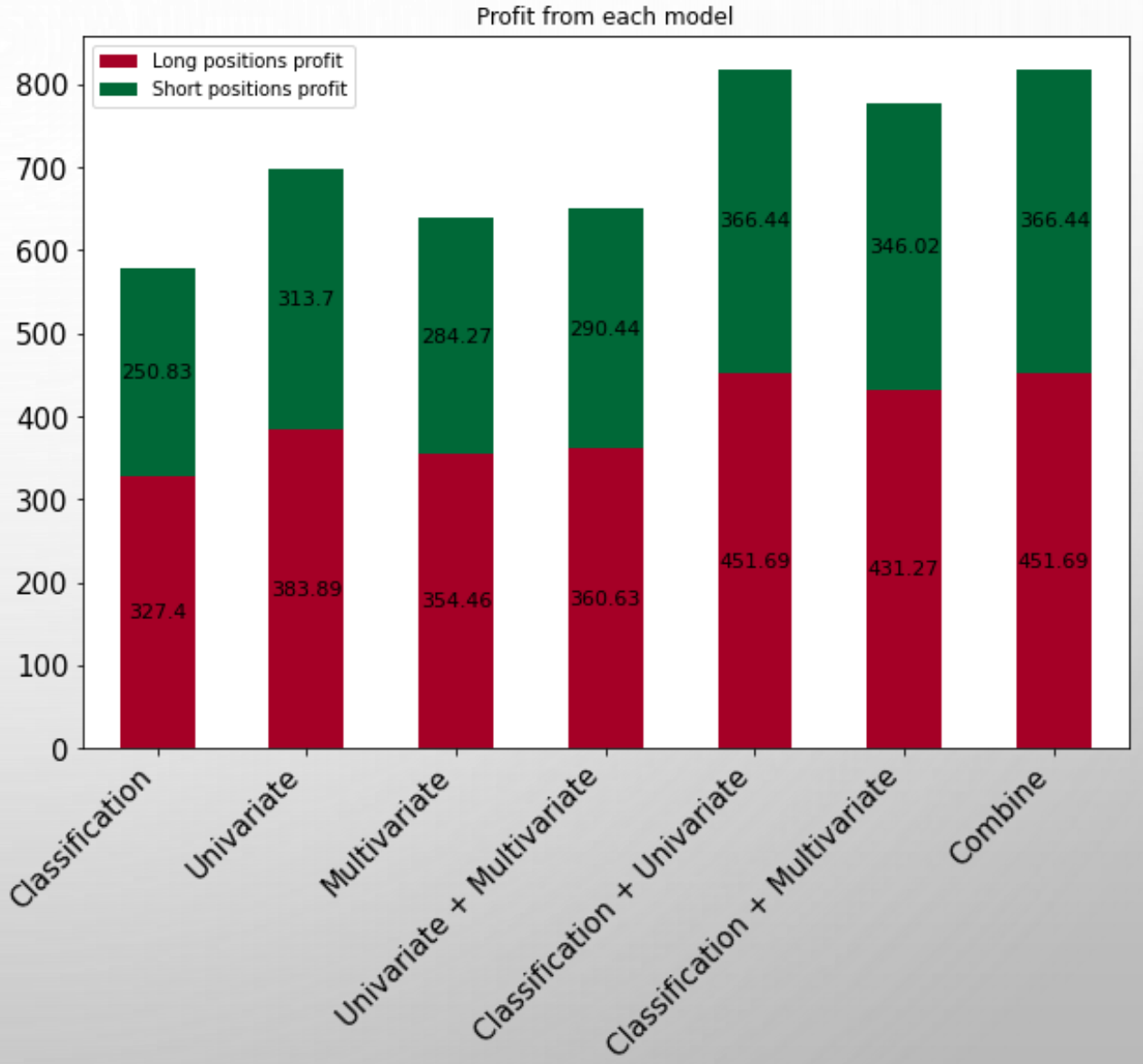






Comparison of different models trading simulations





Thank you

Q & A

https://github.com/MikoPoh/