Can we pick the ‘right’ asset for 2020?

PRASOON KARMACHARYA

Problem Statement

Can we select a stock or a cryptocurrency that will gives us the best ROI at the end of the year?

Criterion:

- Single investment of 100,000 USD on 2020-01-01
- Realize gain based on the adjusted close value of the stock on 2020-12-31

Goal:

- Determine a stock that yields the highest return with high confidence

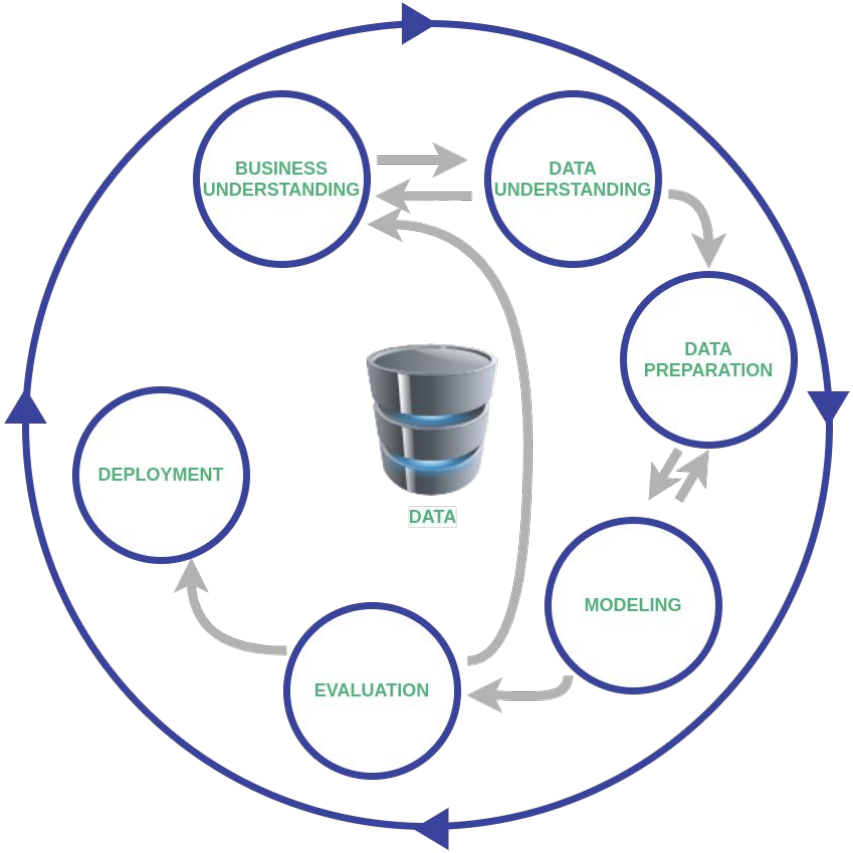
Potential Challenges

- Identifying reliable metrics for stock and crypto asset comparison
- Long term (> 1 month) time-series forecasting of stock is challenging
- Lack of effective and quantifiable exogeneous features that determines the performance of the asset
- Unanticipated events that has huge effects on the stock market

My Approach

- State assumptions to make the problem tractable
- Determine metrics to use for stock comparison:
  - % change stock price
  - Sharpe Ratio
  - Derived metric: θ-Score
- Forecast the stock price for 2020-12-31 using classical models
- Make decision on the stock
- Discuss why stock vs cryptocurrency

Data Science Workflow



Key Assumptions

- Analysis is based on imperfect information (not accounted for exogenous variables such as company's qualitative performance, political/policy influence, pandemic, natural catastrophes etc.)
- Individualized models for each stocks were used instead of generalized model (due to time constraint)

My goal was also to find an unknown, "Dark horse" stock

Metrics - Asset comparison (I)

- **Rate of change** as indicated by the slope of an OLS (regression) trendline
- It signals the growth rate of the stock

$$y_i = \beta_0 + \beta_i \times x_i + \epsilon_i$$

Where,

- β_0 = intercept of linear trend line,
- β_1 = slope of linear trend line
- ϵ_i = error term (aka residual)

- Slope used to determine upward mobility of a given stock

Metrics - Asset comparison (II)

- **Sharpe ratio**: Average return earned in excess of the risk-free rate per unit of volatility.
- It quantifies the relationship between mean of the return and the standard deviation of the return in a given period

$$S_a = \frac{E\left[R_a - R_b\right]}{\sigma_a}$$

S_a = Sharpe Ratio

E = Expected Return

R_a = Asset Return

R_b = Risk Free Return

σ_a = Standard Deviation of excess asset return

- Used to compare relative volatility, higher S_a indicates less volatility

Metrics - Asset comparison (III)

- Derived Metric called **θ-score (β_1,S_a)**

$$\theta\text{-score} = \omega_1 \times \beta_1 + \omega_2 \times \text{sharpe ratio}$$

- ω_1 = weight for the growth rate (β_1)
- ω_2 = weight for the sharpe ratio (S_a)

For, $\omega_1 : \omega_2$ = 0.6:0.4,

$$\theta\text{-score} = 0.6 \times \beta_1 + 0.4 \times \text{sharpe ratio}$$

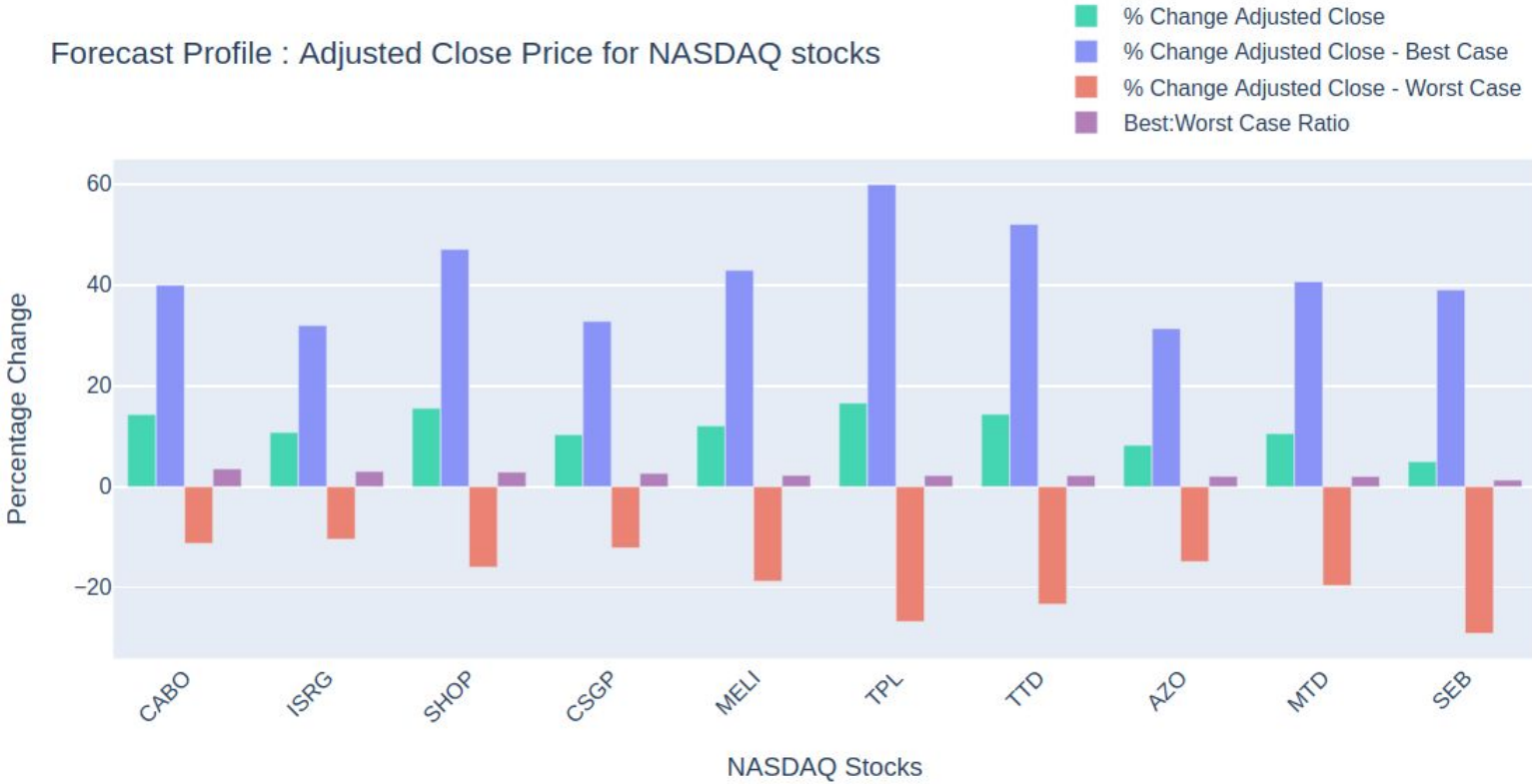
- Used to determine stocks with high return and small volatility
- Narrow down the search space

Top 10 NASDAQ stocks based on θ-score

Adjusted Close Price for the Top 10 NASDAQ stocks

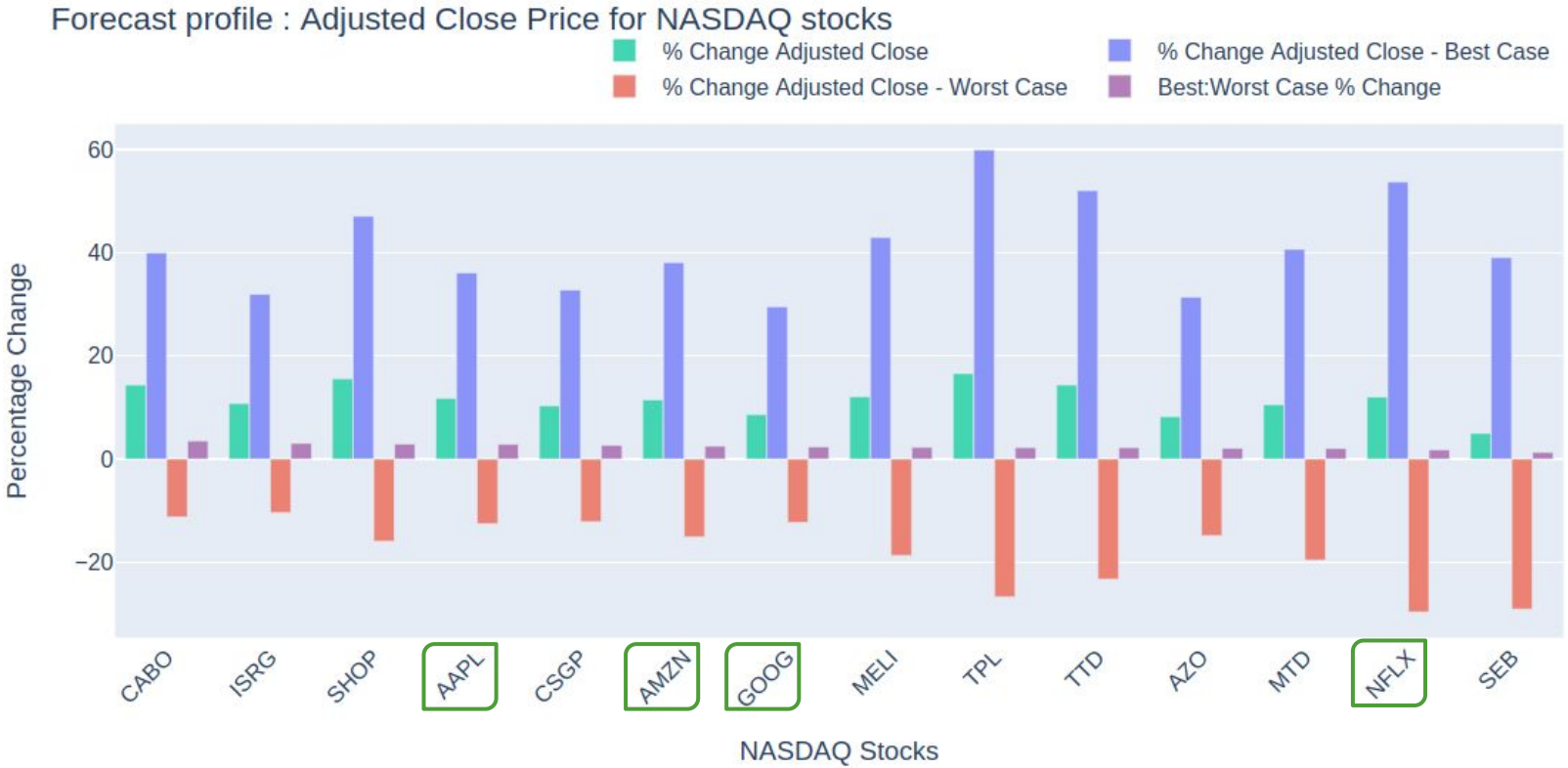
Please see notebook for an interactive plot.

Forecast (SARIMAX) - I

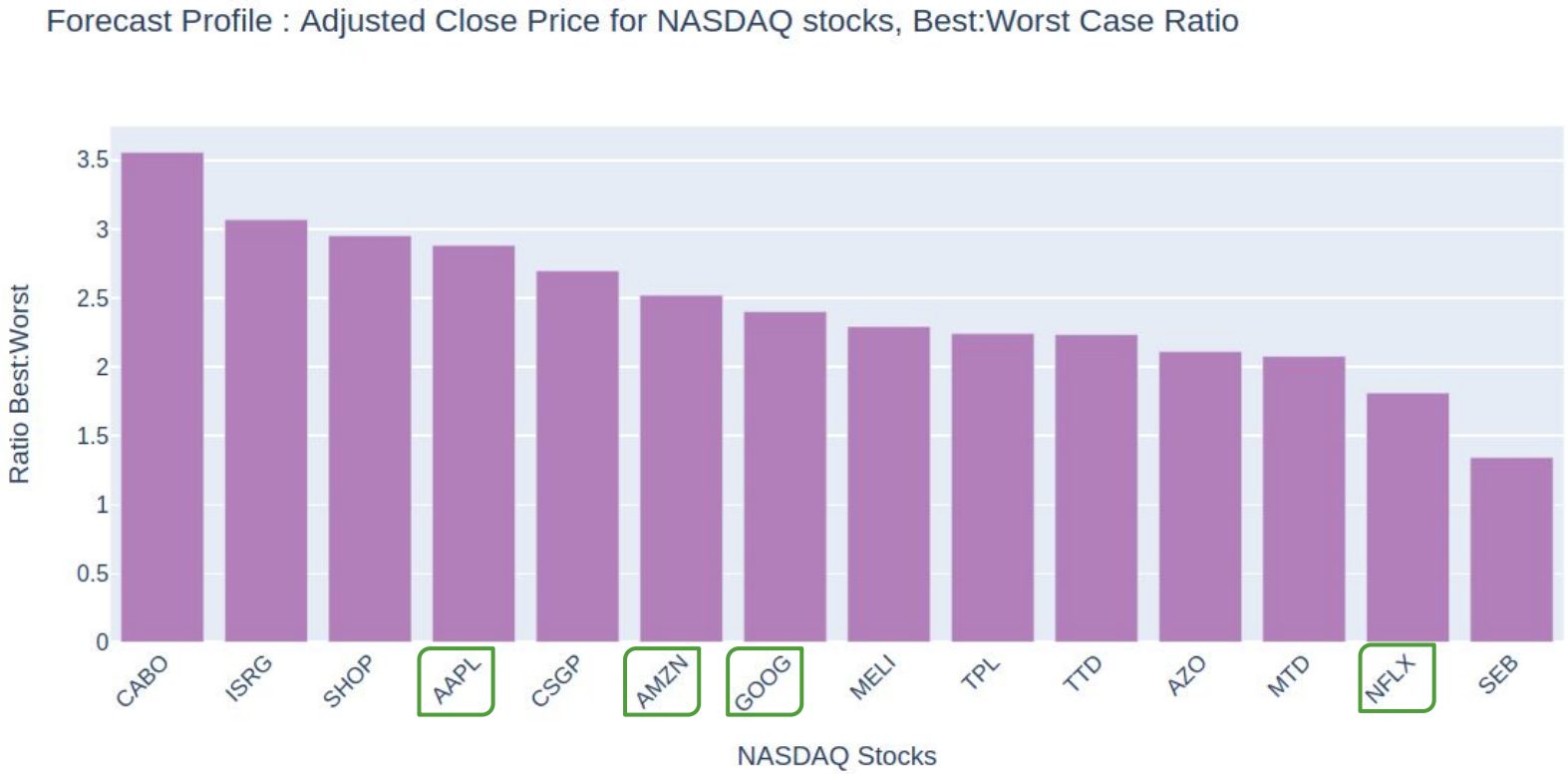


Forecast (SARIMAX) - II

Comparison to the Big Names - I

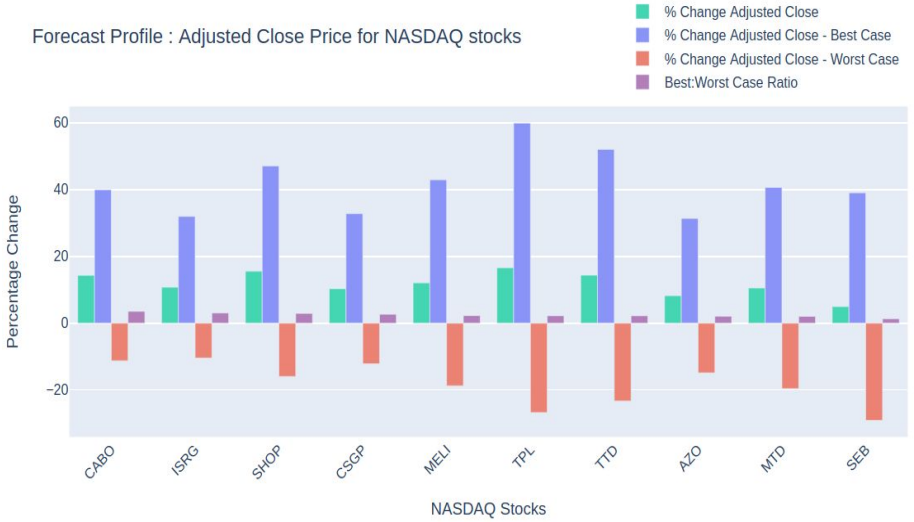


Comparison to the Big Names - II



Favorable Stock

- **CABO** with the highest best:worst ratio of 3.56 as forecasted by the model is the favorable stock
- Corroborated by 2020 data thus-far
- Honorable Mentions: **ISRG**, **SHOP**



Why not Crypto in 2020?

- Speculative Market
- Extremely high volatility
- No significant ties with the economy
- Infant market/asset as far as the time-scale concerning the problem statement;
- Not enough historical data
- No concrete road-map of adoption at least until the end of 2020-12-31 in the US market

Key Takeaways

- Classical time-series modeling fails to forecast the long term stock-price with high accuracy
- However, it is a good tool to use to understand the general direction of the growth of a given stock in a time-scale of one year
- It is almost impossible to build a model that captures catastrophic events

Popular stocks like AMZ, GOOG, AAPL might have better ROI but my goal was to find the “dark-horse” with comparable if not better performance.

Future Considerations

- Include other endogenous variable like (volume, P/E) to strengthen the prediction capacity of SARIMA model
- Include exogenous variables like (stock splits, earnings announcements, trigger news) and use VAR model (multivariate)
- Use NLP to determine the pos/neg sentiment of the news or twitter feed that may affect the stock performance
- Implement RNN models like GRU, LSTM, BiLSTM, LSTM with attention

References

- NASDAQ Listings
- Sharpe Ratio

Tech Stack


pandas


scikit
learn


plotly


TensorFlow


K
Keras


matplotlib


statsmodels

THANK YOU

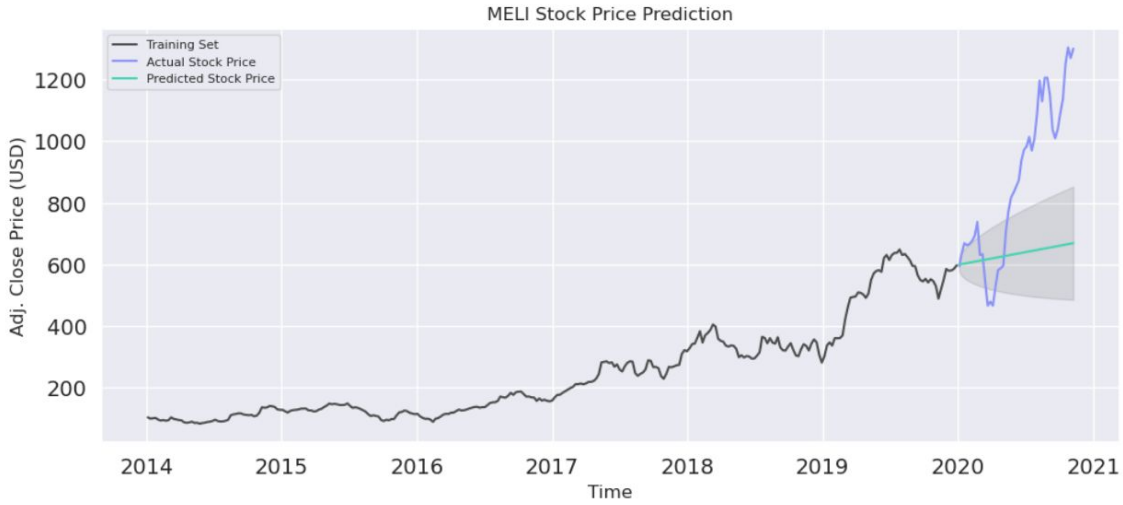
APPENDIX

Stock	SEB
Forecast	4.99 %
(U)pper Bound	39.07 %
(L)ower Bound	-29.09 %
|U:L Ratio|	1.34 %

Stock	CABO
Forecast	14.36 %
(U)pper Bound	39.96 %
(L)ower Bound	-11.23 %
|U:L Ratio|	3.56 %

Stock	TPL
Forecast	16.6 %
(U)pper Bound	59.9 %
(L)ower Bound	-26.7 %
|U:L Ratio|	2.24 %

Stock	TTD
Forecast	14.4 %
(U)pper Bound	52.1 %
(L)ower Bound	-23.3 %
|U:L Ratio|	2.24 %



Stock	MELI
Forecast	12.1 %
(U)pper Bound	42.9 %
(L)ower Bound	-18.7 %
|U:L Ratio|	2.3 %

Stock	SHOP
Forecast	15.6 %
(U)pper Bound	47.1 %
(L)ower Bound	-15.6 %
|U:L Ratio|	3.0 %

Stock	MTD
Forecast	10.5 %
(U)pper Bound	40.7 %
(L)ower Bound	-19.6 %
|U:L Ratio|	2.1 %

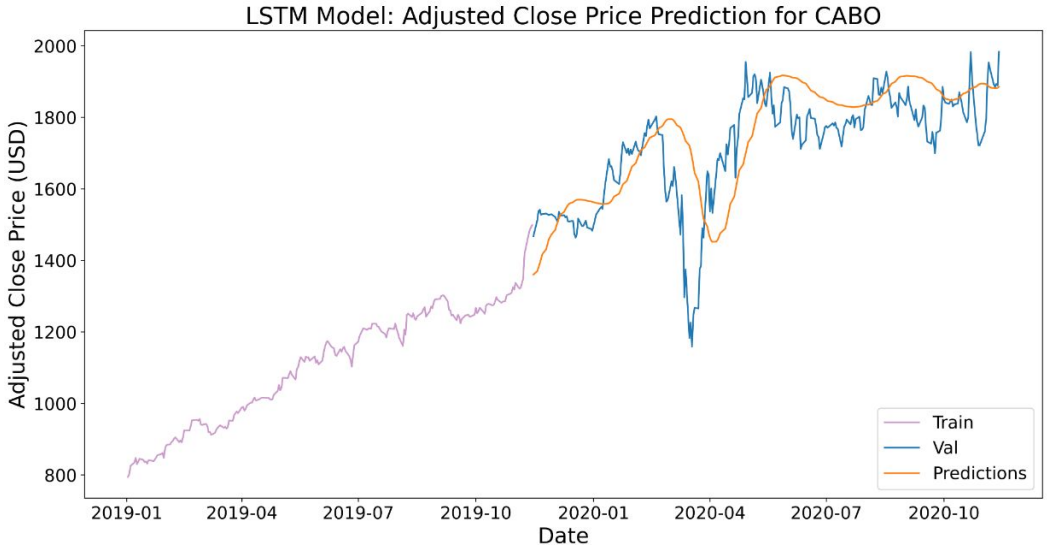
Stock	ISRG
Forecast	10.8 %
(U)pper Bound	32.0 %
(L)ower Bound	-10.4 %
|U:L Ratio|	3.1 %

Stock	AZO
Forecast	8.3 %
(U)pper Bound	31.4 %
(L)ower Bound	-14.5 %
|U:L Ratio|	2.1 %

Stock	CSGP
Forecast	10.3 %
(U)pper Bound	32.8 %
(L)ower Bound	-12.2 %
|U:L Ratio|	2.7 %

LSTM model prediction

- Misleading since this implementation use adj. close price of previous 60 days and we do not have the data for 2020
- Really good for short-term forecasting (RMSE ~ 100 USD)



Deliverables

- The winning stock or cryptocurrency, along with some mathematical representation of how confident you are.
- A notebook containing your work
- A repo containing the notebook and any other relevant information
- At least 4 different types of graphs to visualize your approach
- Optional: a quick way to track the current rank of your choice for a given day of 2020.