Time Series Analysis Model Evaluation

Summary

Goal: Evaluate several machine learning models to utilize time series data and data on financial fundamentals to determine the most profitable machine learning model.

Method: Choose three types of models, train and test data, evaluate model performance.

Hypothesis: Our group predicts the Recurrent Neural Network model will reveal the most profitable performance.

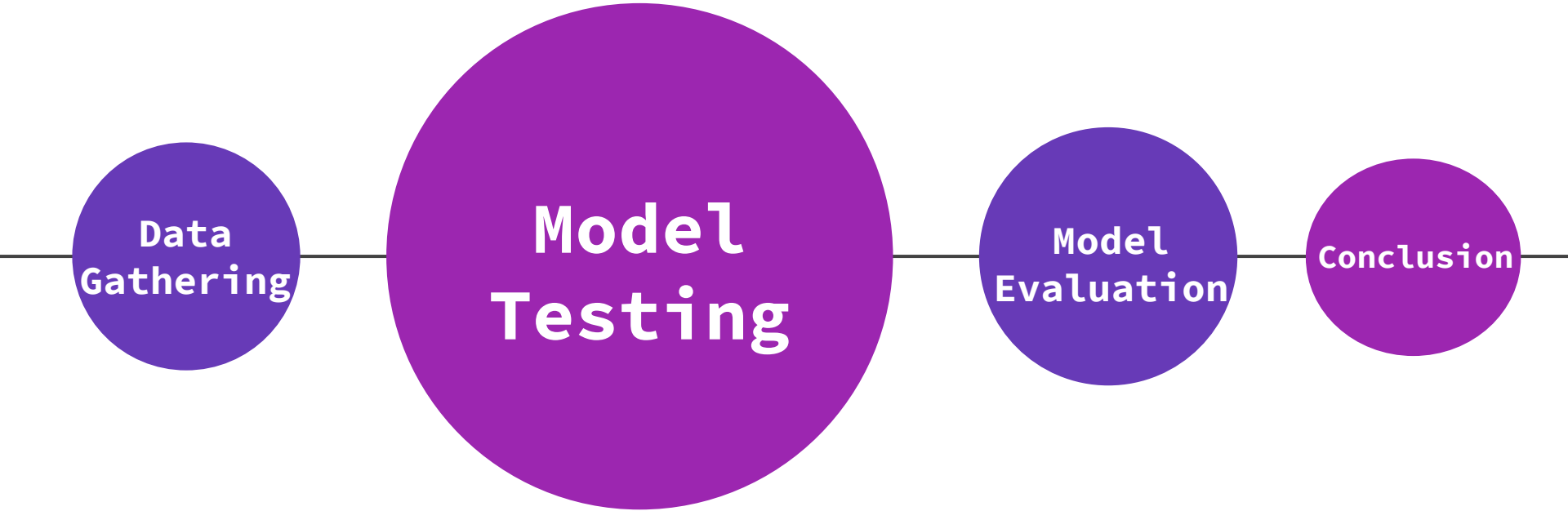
Research Questions

1. Will our portfolio performance reveal financial fundamental metrics can be used to predict stock performance?

2. Using business fundamental financial statement data including income statements, show stock price performance?

3. Can we predict a percentage return range based on our model performance?

Timeline of Events



Data Gathering

- Google Finance - Google Sheets - Historical Price Data
- Financial Modeling Prep - Income Statement Data
- Quandl API for Economic Data: GDP, 10-Year Treasury rate, Industrial production index, Consumer Price Index,
- Yahoo Finance - Historical Price Data

Model Testing

- Initially used stock close prices for LSTM model and evaluated our results

- Chose second model, ARIMA, and used stock close prices and evaluated our results

- Implemented third “type” of model - RNN LSTM Multivariate - used close price as well as income statement data gathered and combined into one data frame

RNN LSTM

- MSE error was extremely high for certain stocks, but not all
- Error convergence would not occur
- Companies that had high MSE error values were two tech companies, and have also been publically traded for the least amount of time

ARIMA

- Large MSE error as well on the ARIMA model
- Ran both model tests on close price data to evaluate all of the chosen stocks to see results

Univariate LSTM - Overview

Data

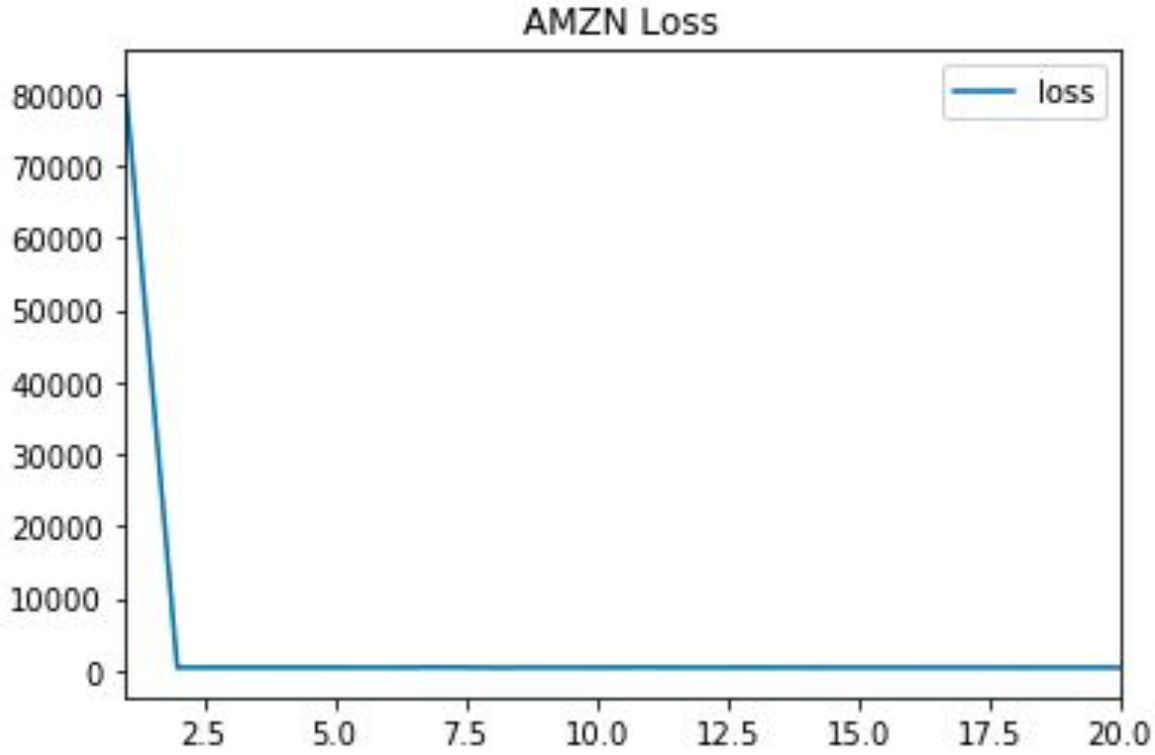
n_steps	3
n_features	1

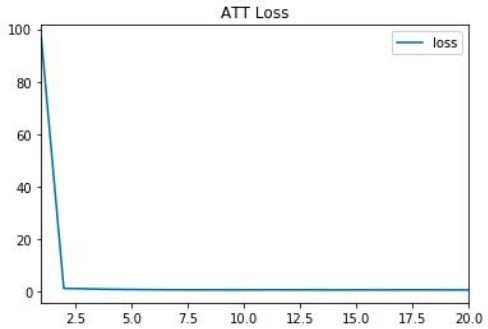
Model Parameters

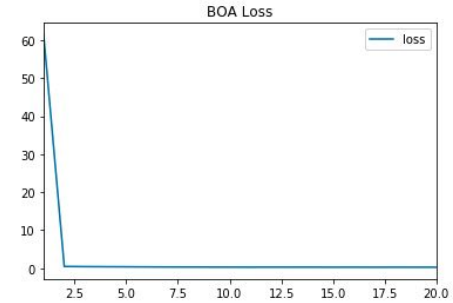
LSTM Units	50
Activation	‘relu’
Epochs	20
Neurons	1
Optimizer	‘adam’

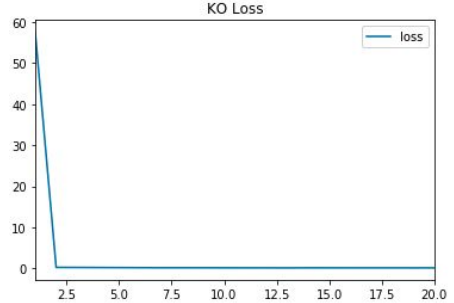
Univariate LSTM - MSE

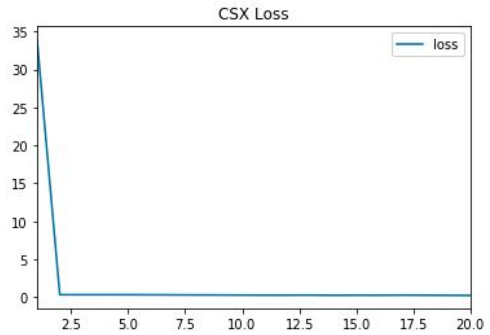
AMZN	346.878
ATT	0.7219
BOA	0.295
KO	0.1883
CSX	0.3035
JPM	1.0175
ODFL	1.199
PG	0.6363
TSLA	24.669
VZ	0.3765

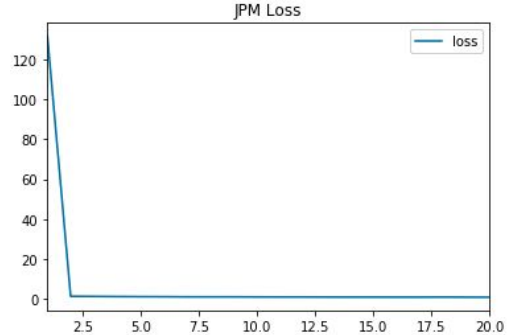


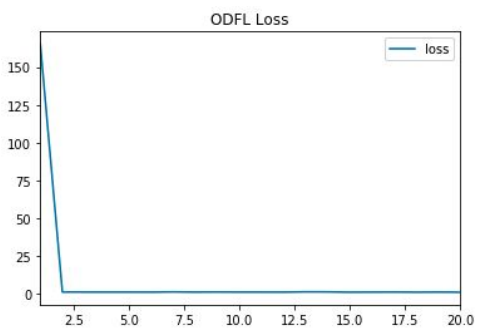

ATT Loss
loss

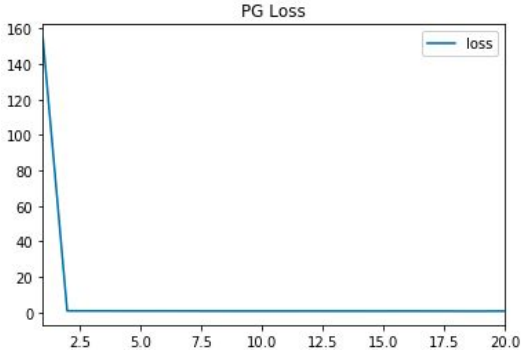

BOA Loss
loss

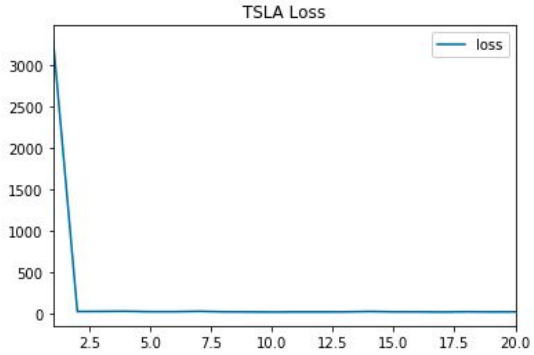

KO Loss
loss

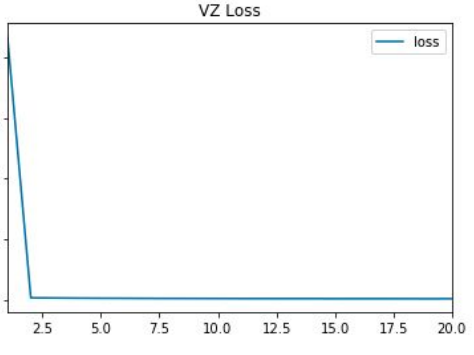

CSX Loss
loss


JPM Loss
loss

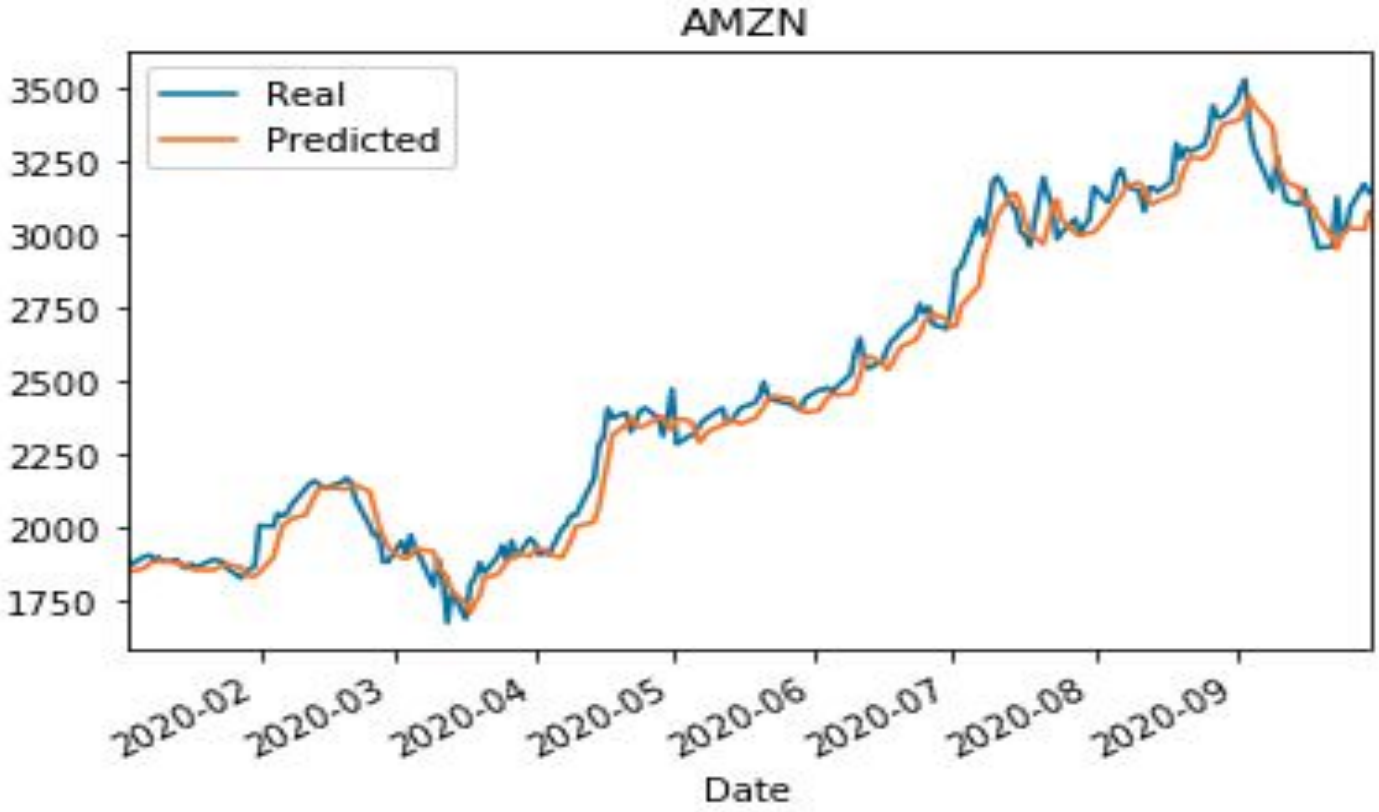

ODFL Loss
loss

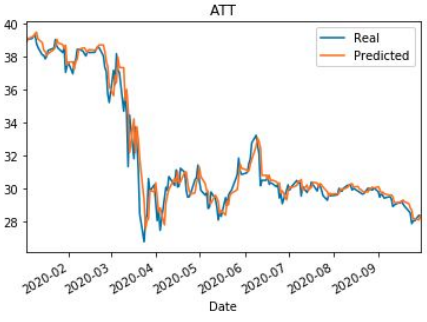

PG Loss
loss

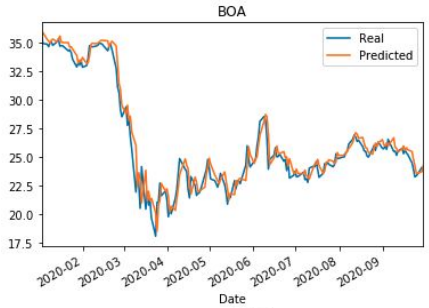

TSLA Loss
loss

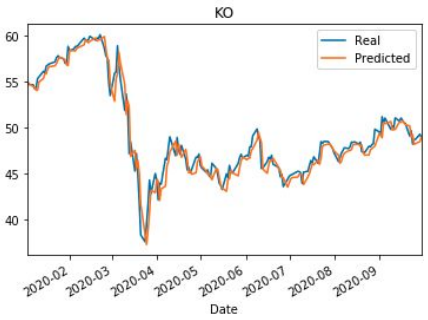

VZ Loss
loss

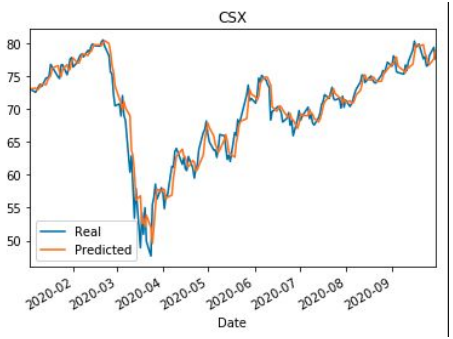
Univariate LSTM - Actual vs. Predicted Prices

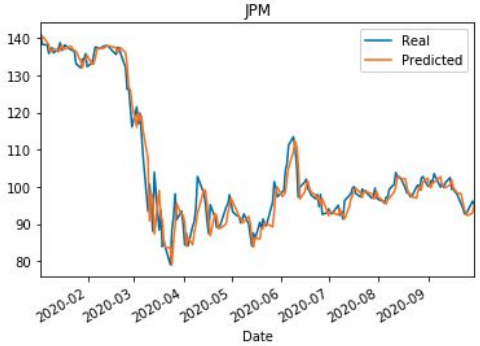


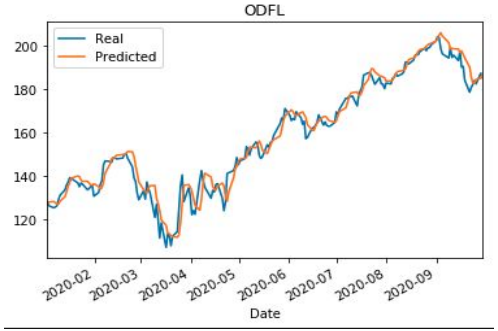

ATT
Real
Predicted
Date

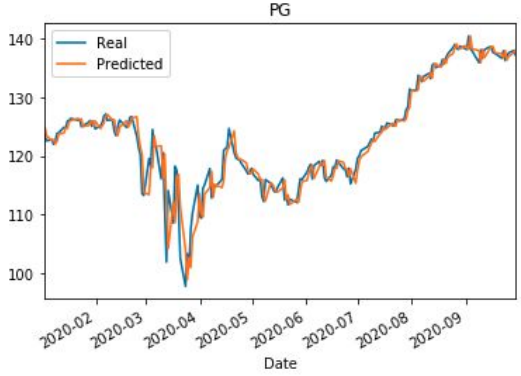

BOA
Real
Predicted
Date

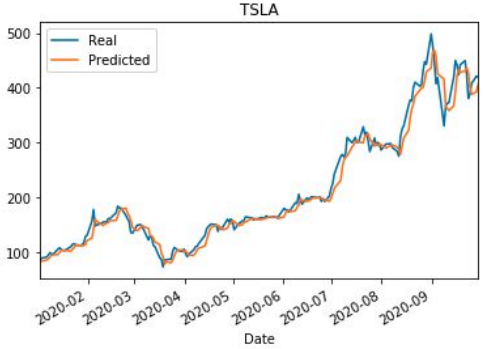

KO
Real
Predicted
Date

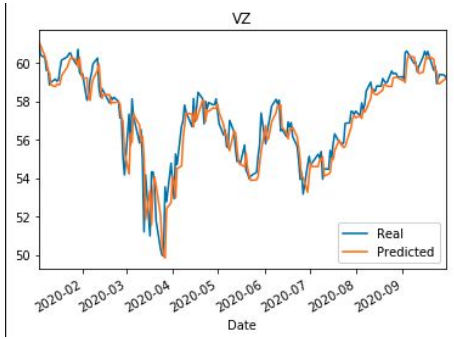

CSX
Real
Predicted
Date


JPM
Real
Predicted
Date


ODFL
Real
Predicted
Date


PG
Real
Predicted
Date


TSLA
Real
Predicted
Date


VZ
Real
Predicted
Date

AutoRegressive Integrated Moving Average (ARIMA) - Overview

Data

Train Data	70%
Test Data	30%

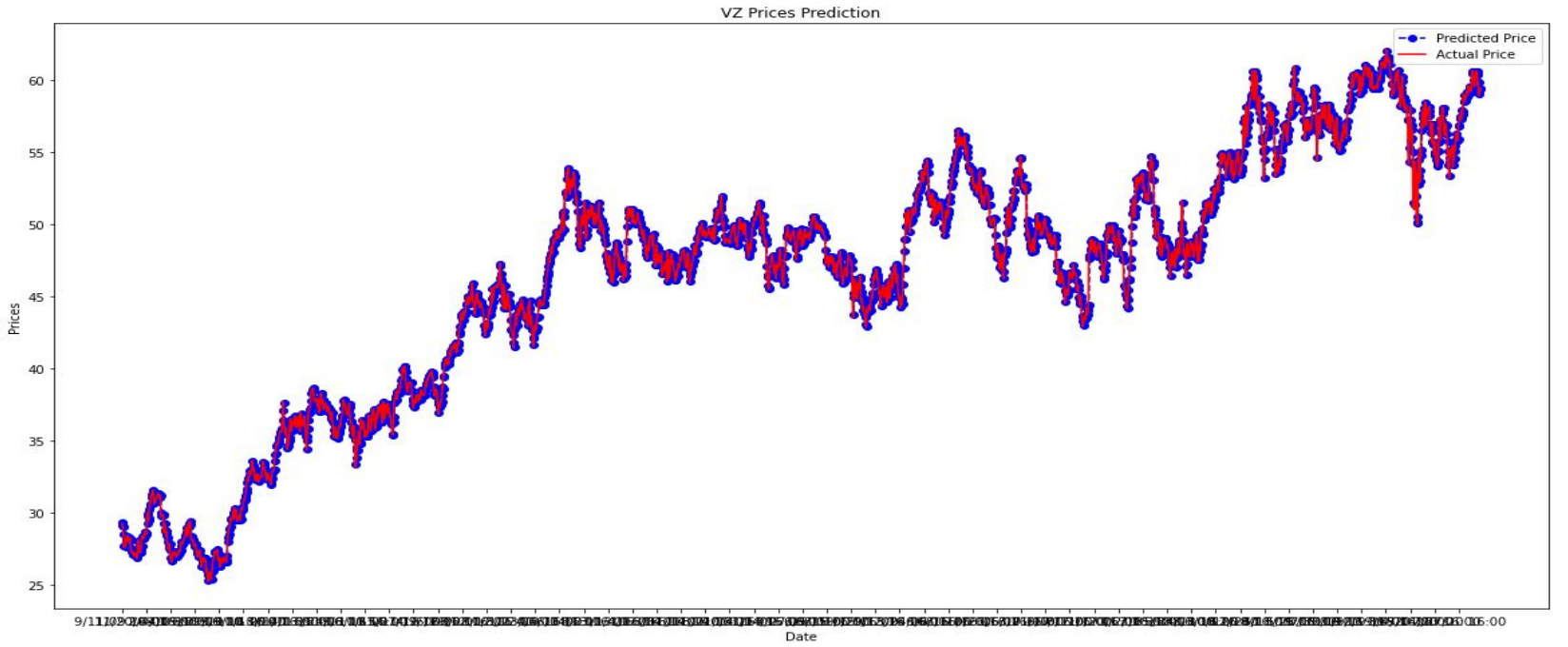
Arima Parameters

p:the number of lag observations in the model; also known as the lag order.	4
d:the number of times that the raw observations are differenced; also known as the degree of differencing.	1
q:the size of the moving average window; also known as the order of the moving average.	0

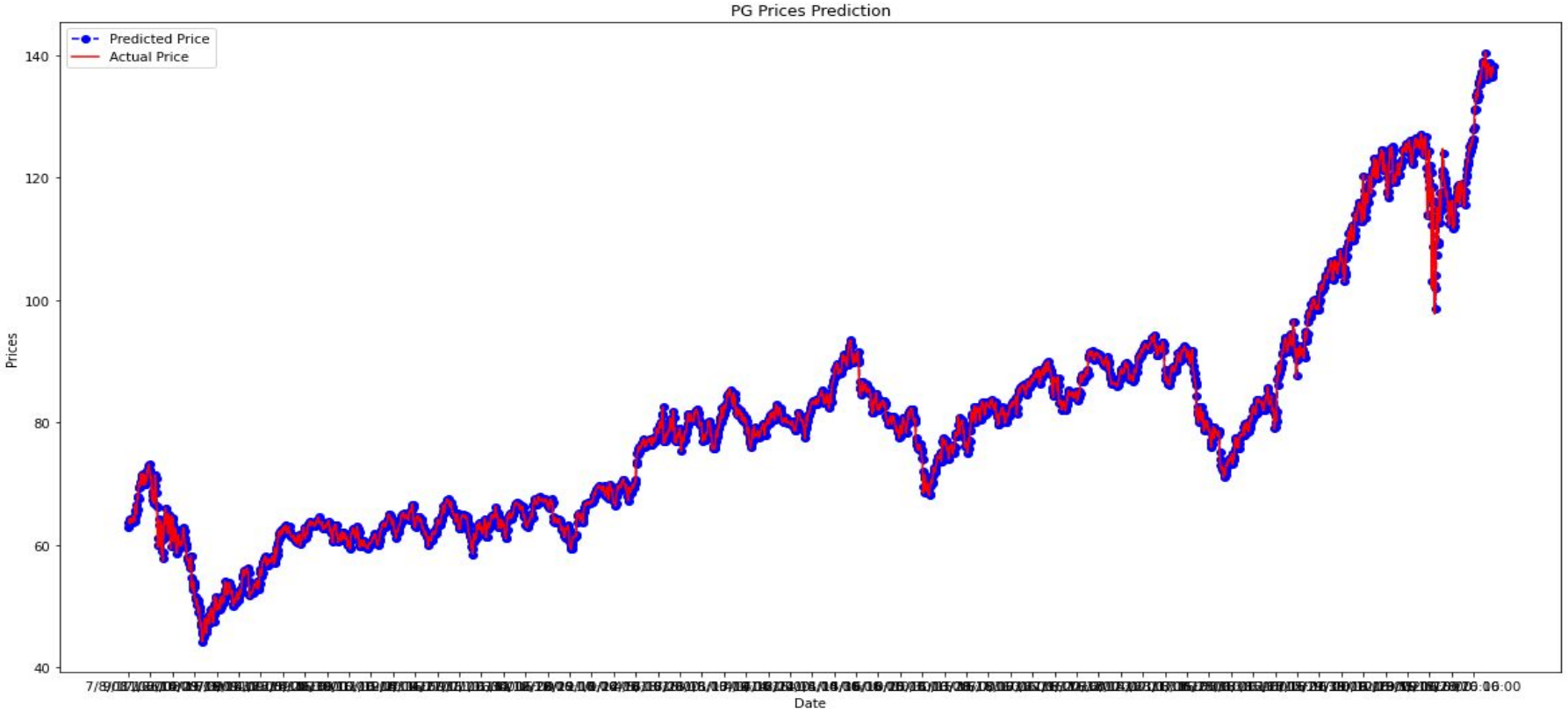
ARIMA Mean Squared Error(MSE)

- AMZN:899.1964
- T:0.1619
- BOA:0.3509
- COKE:0.2716
- CSX:0.60347
- JPM:1.9460
- ODFL:2.3335
- PG:1.0331
- TSLA:60.5745
- VZ:0.2851

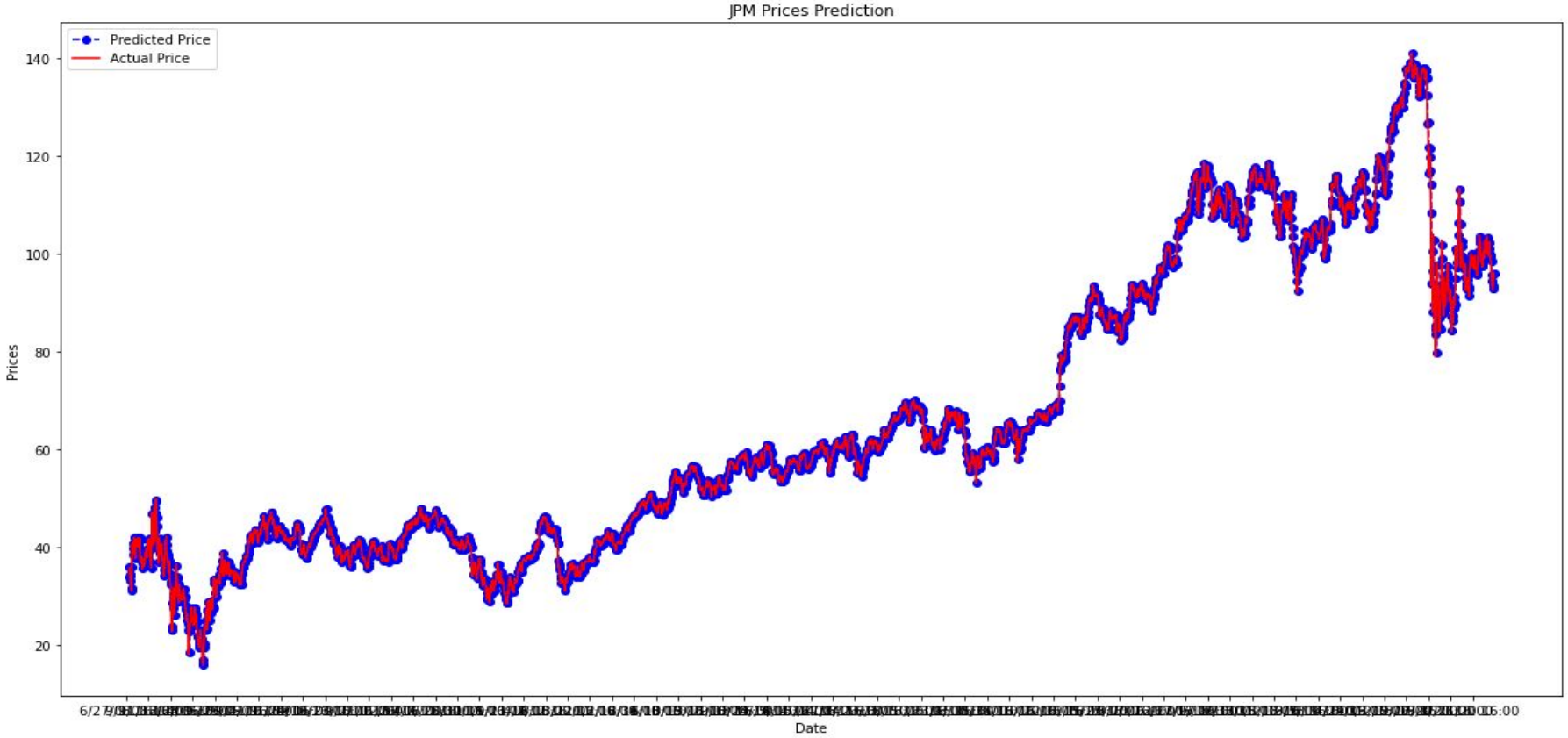
Some ARIMA Actual vs. Predicted Prices



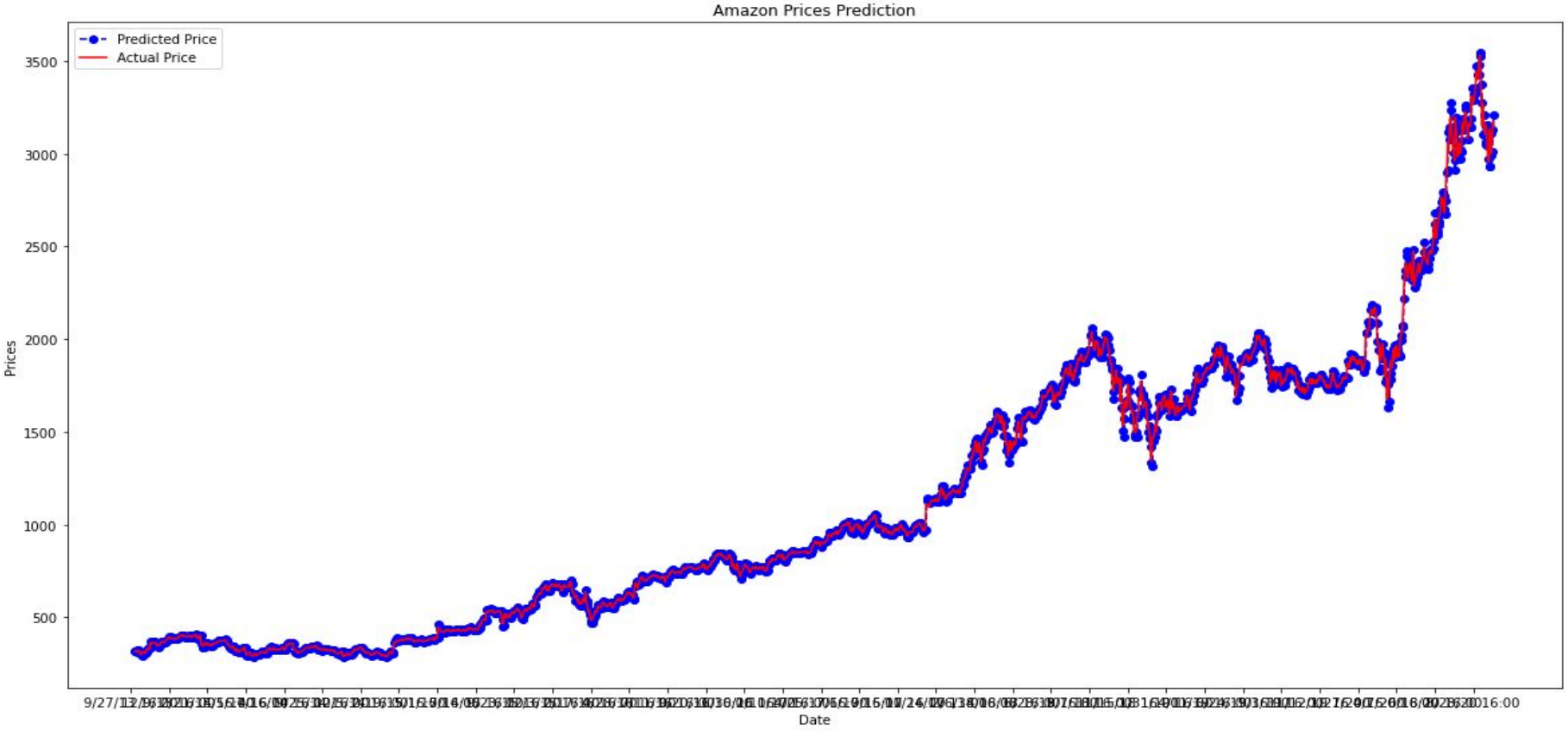
PG Prices Prediction



JPM Prices Prediction



Amazon Prices Prediction



RNN LSTM Multivariate Model - Overview

Model = Sequential

LSTM Unites = 50

Dropout = 0.2

Dense Units = 1

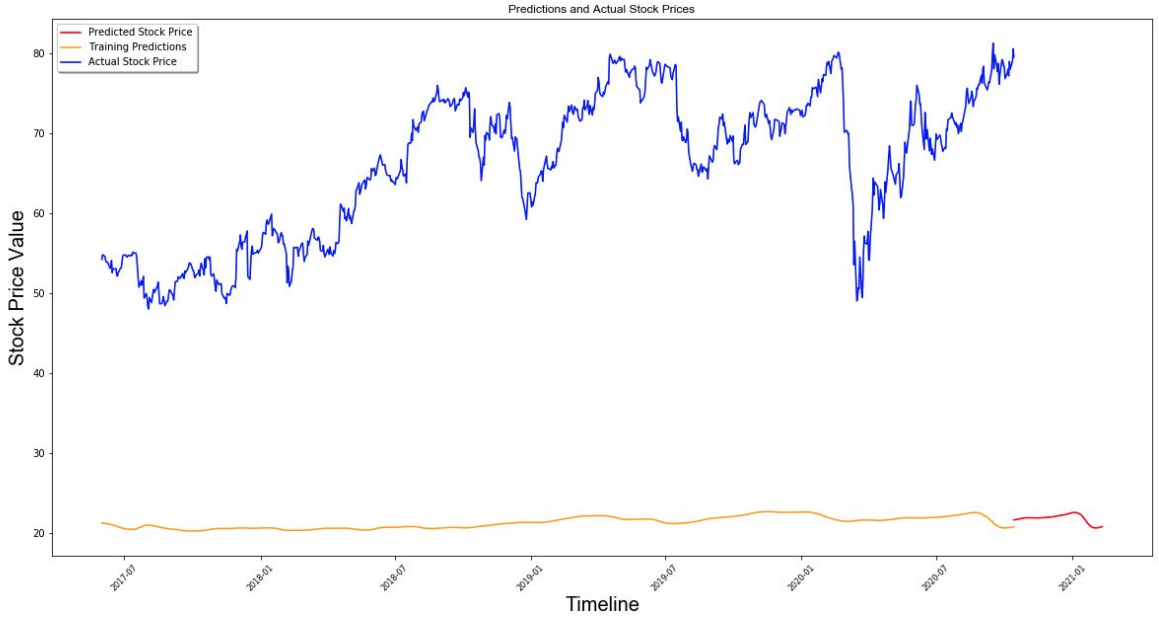
Optimizer = Adam

Loss = Mean Squared Error

Features (parameters):

Open	High
Low	Close
Adj Close	Volume

RNN LSTM Multivariate Model - Results



Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 730, 50)	11000
dropout (Dropout)	(None, 730, 50)	0
lstm_1 (LSTM)	(None, 730, 50)	20200
dropout_1 (Dropout)	(None, 730, 50)	0
lstm_2 (LSTM)	(None, 730, 50)	20200
dropout_2 (Dropout)	(None, 730, 50)	0
lstm_3 (LSTM)	(None, 50)	20200
dropout_3 (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51

Total params: 71,651
Trainable params: 71,651
Non-trainable params: 0

RNN LSTM Multivariate Model - Overview

Model = Sequential

LSTM Units = 50

Dropout = 0.2

Dense Units = 1

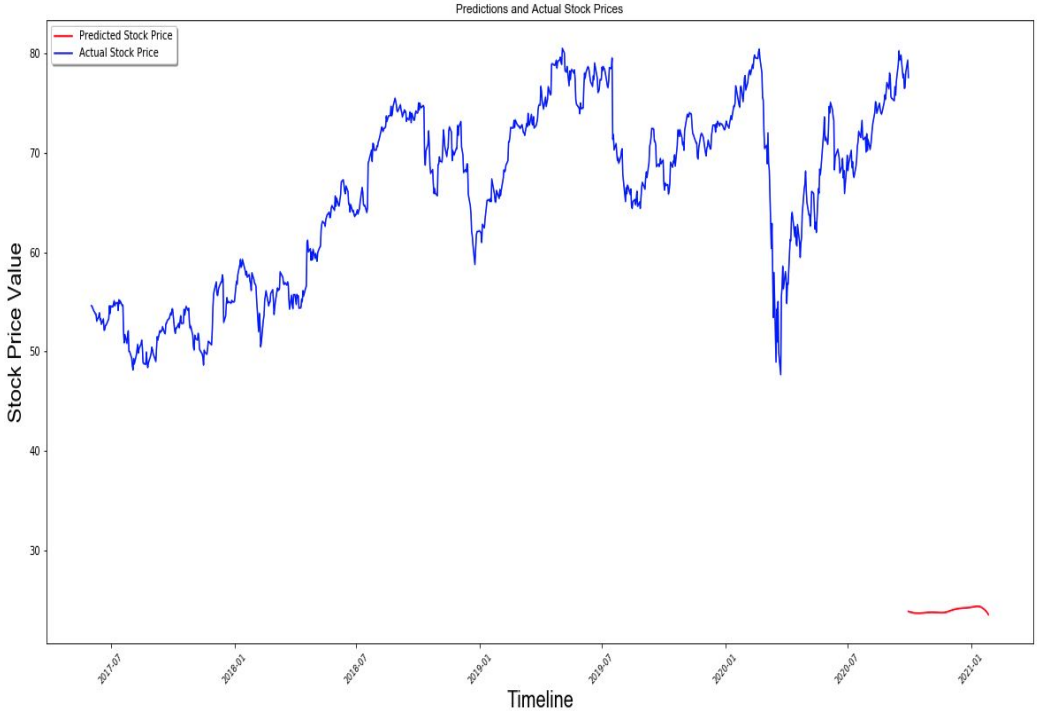
Optimizer = Adam

Loss = Mean Squared Error

Features (parameters):

Close	Revenue	Cost Of Revenue	Gross Profit
Gross Profit Ratio	Cost and Expenses	Interest Expense	Depreciation & amortization
EBITDA	EBITDA Ratio	Operating Income	Operating Income Ratio
Net Income	Earnings Per Share	Net Income Ratio	Income Before Tax

RNN LSTM Multivariate Model - Results



Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 730, 50)	13800
dropout (Dropout)	(None, 730, 50)	0
lstm_1 (LSTM)	(None, 730, 50)	20200
dropout_1 (Dropout)	(None, 730, 50)	0
lstm_2 (LSTM)	(None, 730, 50)	20200
dropout_2 (Dropout)	(None, 730, 50)	0
lstm_3 (LSTM)	(None, 50)	20200
dropout_3 (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51

Total params: 74,451
Trainable params: 74,451
Non-trainable params: 0

Model Evaluation

LSTM Univariate

- Model accuracies for all chosen stocks were “zero”
- Would not recommend model

ARIMA

- Training and Testing Data were split 70/30, but the model lacked sufficient predictive value

LSTM Multivariate Model

- Adding features decreased Mean Squared Error, but did not add predictive value for Close stock prices

Concluding Remarks

Names:
Bryan Castillo
Andrew Smith-Mui
Yao Kouassi