Capstone Project

Stock Price Prediction with ARIMAX, SARIMAX, GRU, LSTM

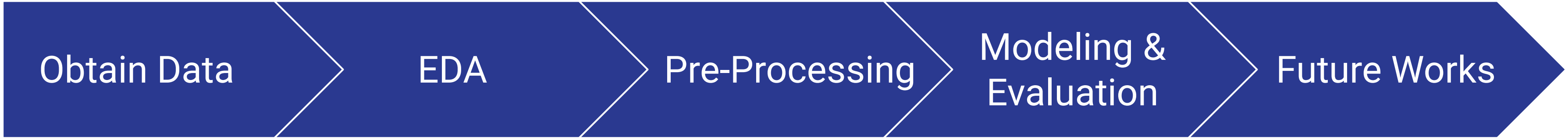
Henri Oei - DSIF2

The problem

The objective is to create a predictive model to forecast future stock closing prices.

For the purpose of this presentation I will be showcasing the AMD stock ticker.

Overview



Obtain Data

EDA

Pre-Processing

Modeling & Evaluation

Future Works

Pandas Datareader

- Tiingo
- IEX
- Alpha Vantage
- Econdb
- Enigma
- Quandl
- St.Louis FED (FRED)
- Kenneth French’s data library
- World Bank
- OECD
- Eurostat
- Thrift Savings Plan
- Nasdaq Trader symbol definitions
- Stooq
- MOEX
- Naver Finance
- Yahoo Finance

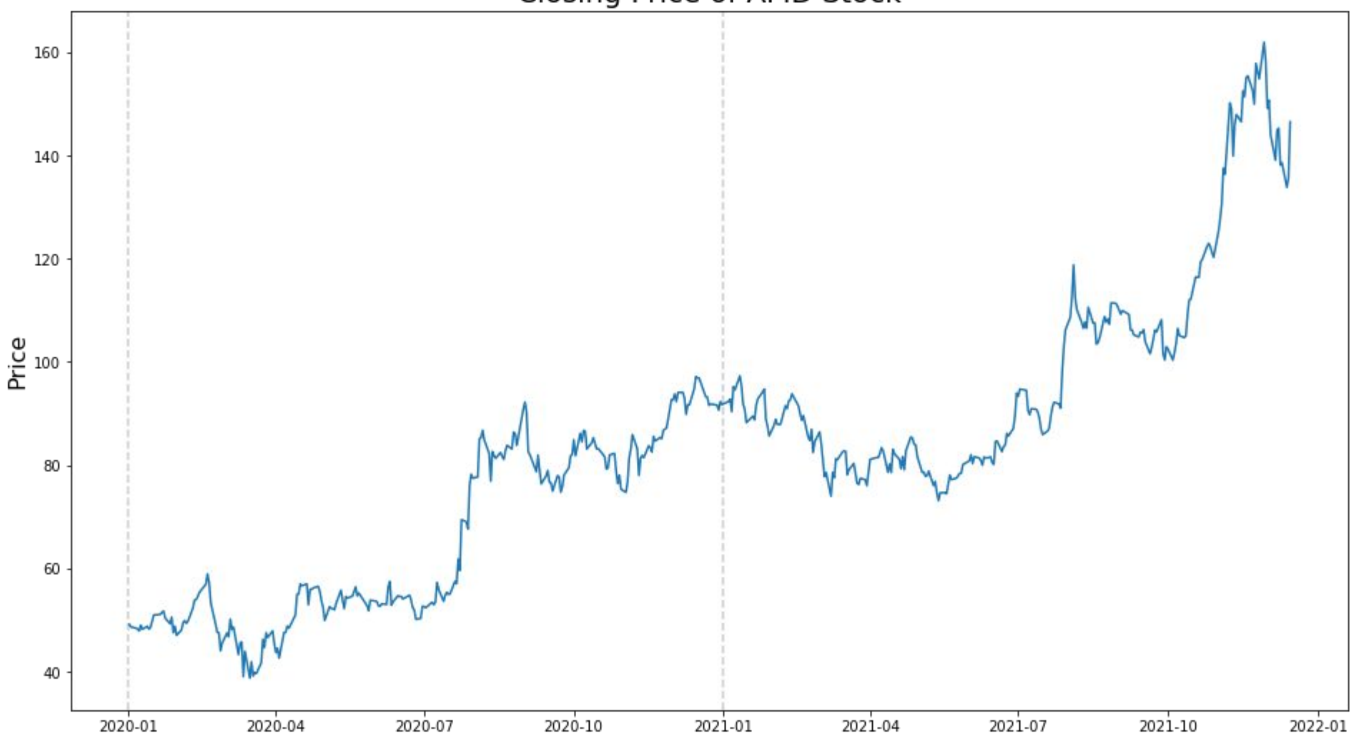
Obtain Data

EDA

Pre-Processing

Modeling & Evaluation

Future Works



Closing Price of AMD Stock

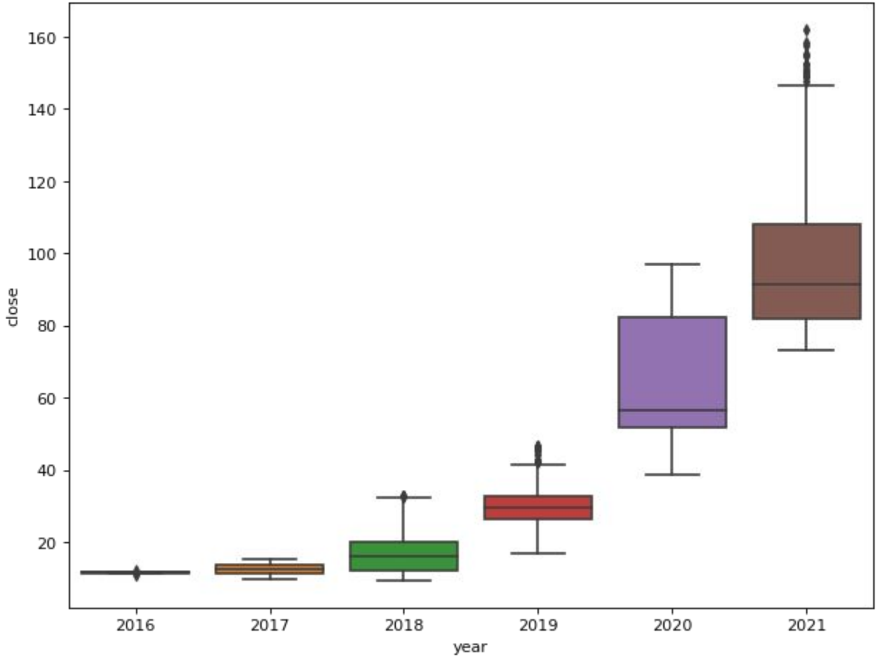
Obtain Data

EDA

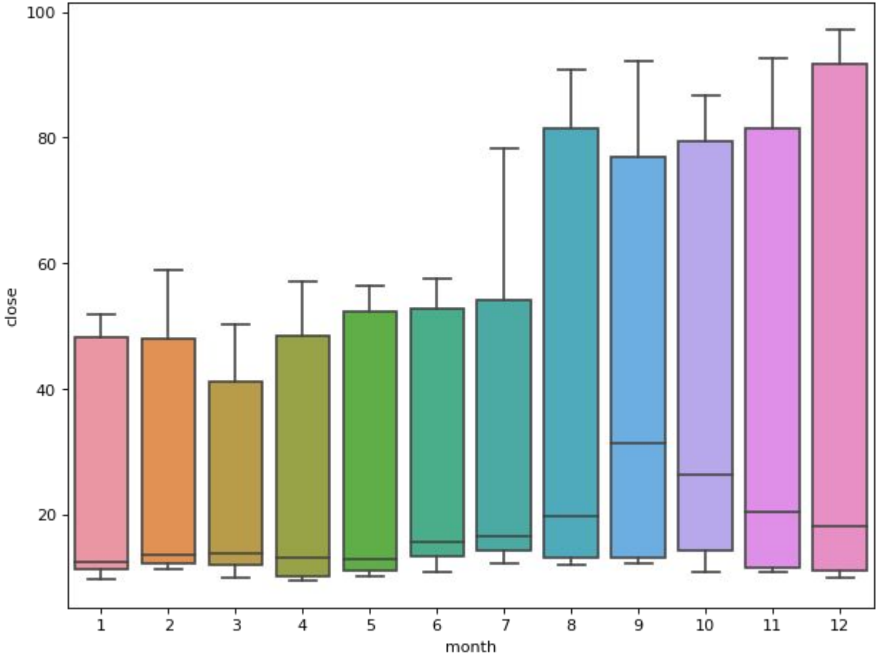
Pre-Processing

Modeling & Evaluation

Future Works



Year-wise Box Plot
(The Trend)



Month-wise Box Plot
(The Seasonality)

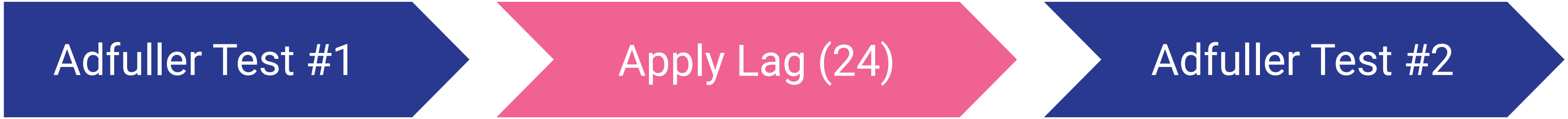
Obtain Data

EDA

Pre-Processing

Modeling & Evaluation

Future Works



Adfuller Test #1

Apply Lag (24)

Adfuller Test #2

ADF Test Statistic :

1.3842659609766095

p-value : 0.997041479296224

Number of Observations Used : 1237

Comment : Time series is non-stationary

Procedure

Day #1 to Day #24 = N/A

Day #25 = Day #25 - Day #1

Day #26 = Day #26 - Day #2

Day #27 = Day #27 - Day #3

.

.

ADF Test Statistic :

-8.610593374916578

p-value : 6.491890430629663e-14

Number of Observations Used : 1209

Comment : Time series is stationary

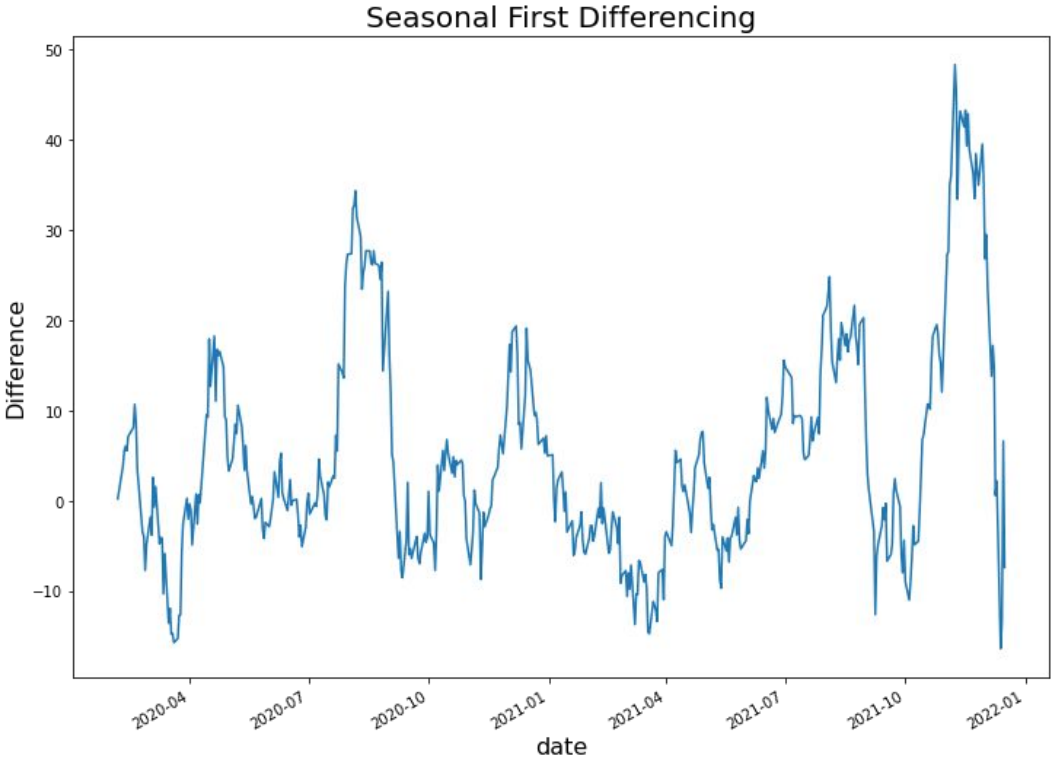
Obtain Data

EDA

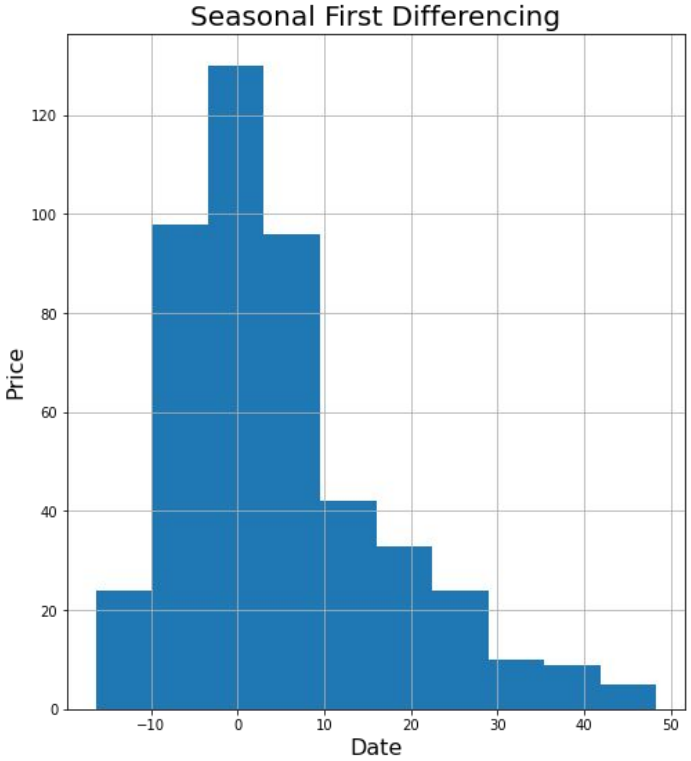
Pre-Processing

Modeling & Evaluation

Future Works



Seasonal First Differencing



Seasonal First Differencing

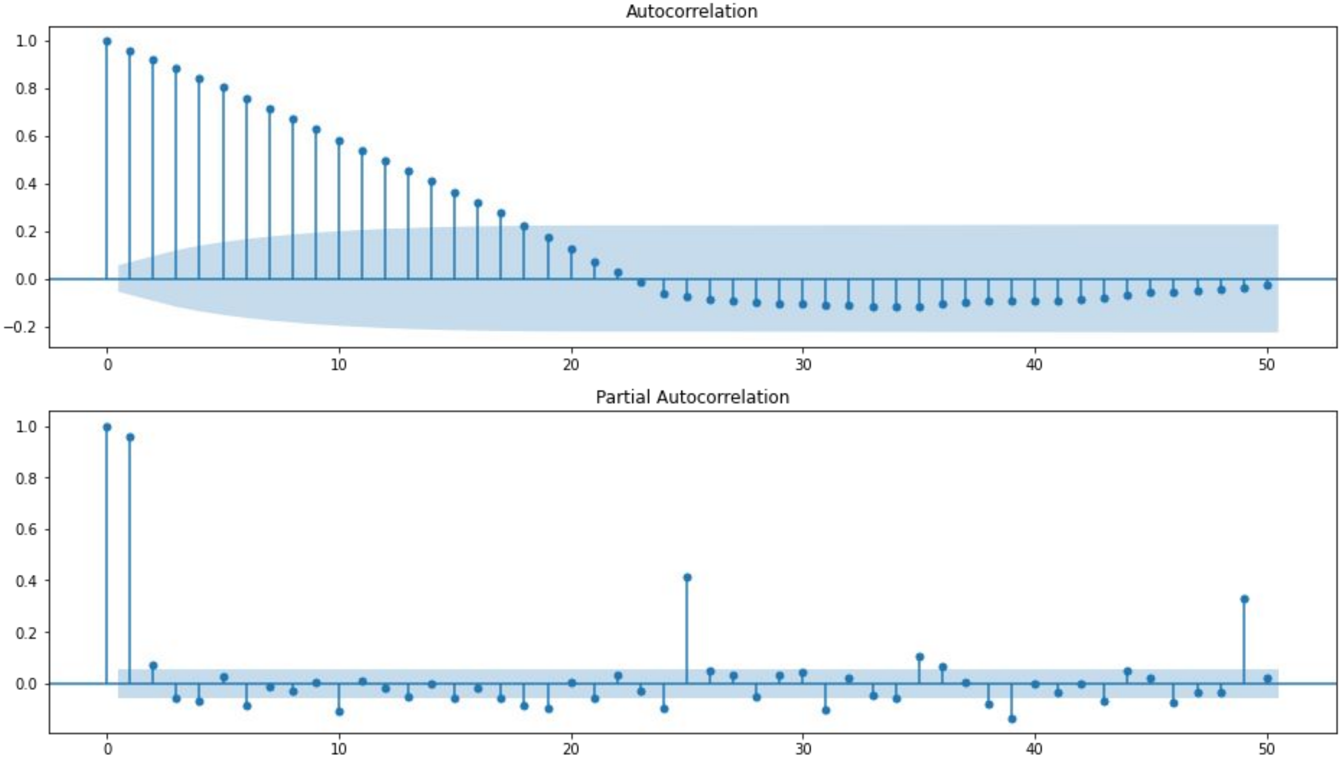
Obtain Data

EDA

Pre-Processing

Modeling & Evaluation

Future Works



Obtain Data

EDA

Pre-Processing

Modeling & Evaluation

Future Works

ARIMAX Architecture

80 - 20 Train Test Split

Order = (1, 1, 1), # (p, d, q)

Exog = Lag(24)

p : Trend Autoregression Order
d : Trend difference order
q : Trend moving average order
m : The number of time steps for a single seasonal period
Exog: OLS Regressor

SARIMAX Architecture

80 - 20 Train Test Split

Order = (1, 1, 1), # (p, d, q)

Seasonal Order = (1, 1, 1, 24), # (p, d, q, m)

Exog = Lag(24)

GRU Architecture

80 - 20 Train Test Split

Layer #1 - GRU

Units : 128
Activation : Tanh
Recurrent Activation : Sigmoid

Layer #2 - GRU

Units : 64
Activation : Tanh
Recurrent Activation : Sigmoid

Layer #3 - Dropout

Rate : 0.2

Layer #4 - Dense

Units : 1

Evaluator

Loss : Mean Squared Error
Optimizer : Adam

LSTM Architecture

80 - 20 Train Test Split

Layer #1 - LSTM

Units : 128
Activation : Tanh
Recurrent Activation : Sigmoid

Layer #2 - LSTM

Units : 64
Activation : Tanh
Recurrent Activation : Sigmoid

Layer #3 - Dropout

Rate : 0.2

Layer #4 - Dense

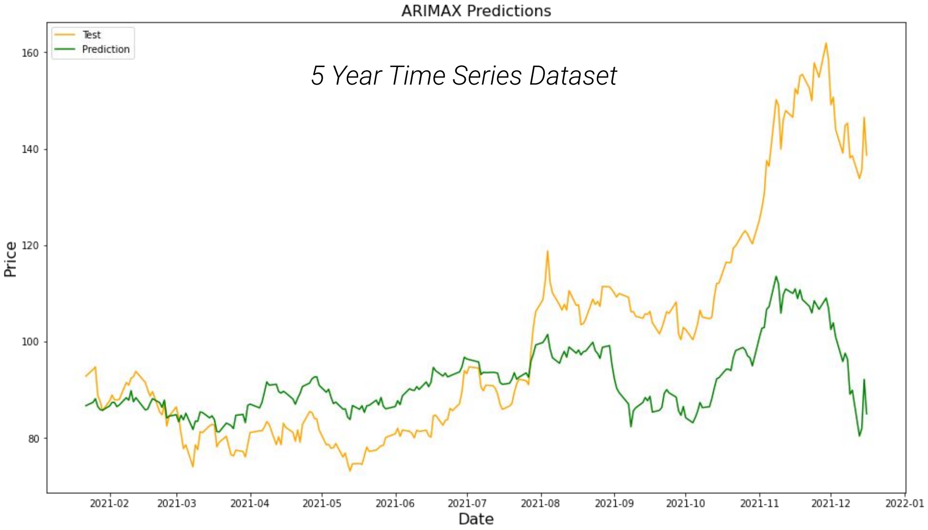
Units : 1

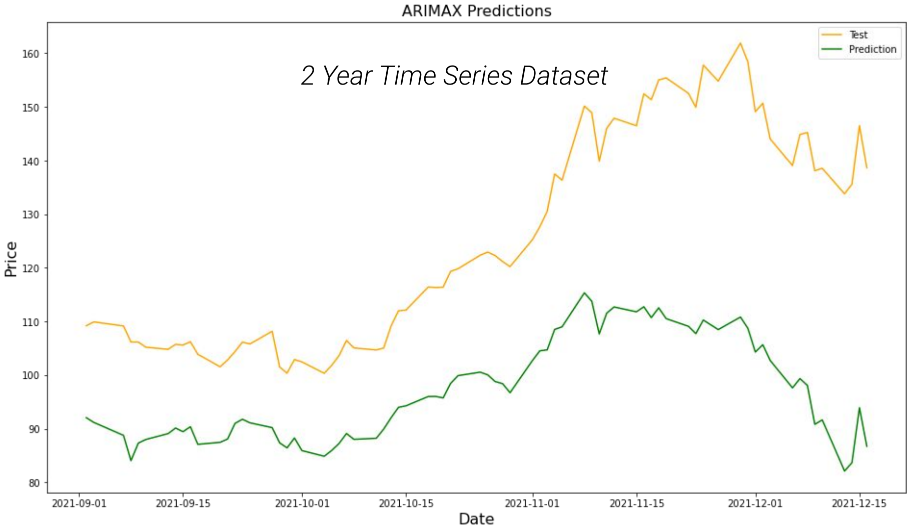
Evaluator

Loss : Mean Squared Error
Optimizer : Adam

ARIMAX Model

80 - 20 Train Test Split





Obtain Data

EDA

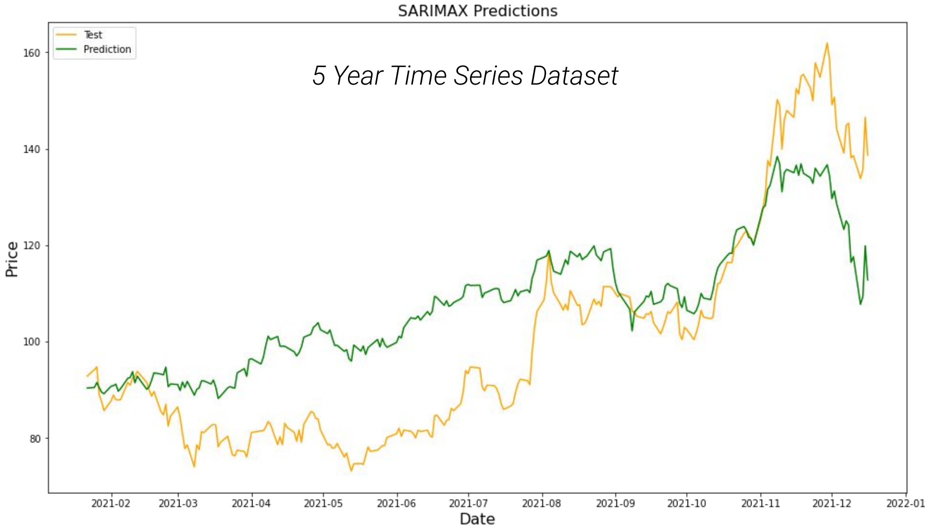
Pre-Processing

Modeling & Evaluation

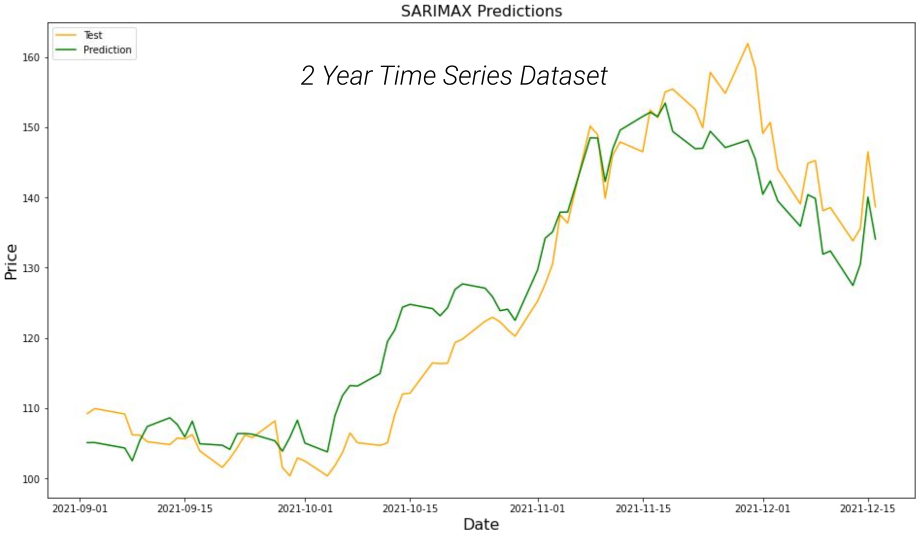
Future Works

SARIMAX Model

80 - 20 Train Test Split


SARIMAX Predictions

5 Year Time Series Dataset

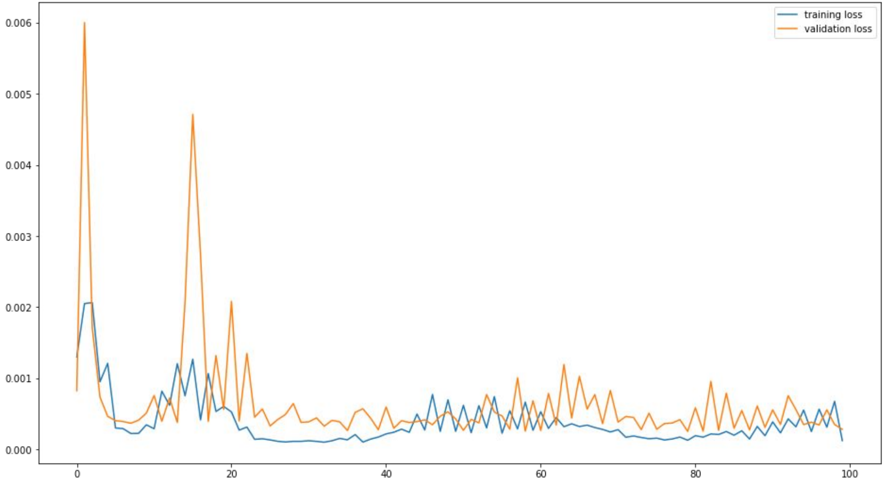

SARIMAX Predictions

2 Year Time Series Dataset

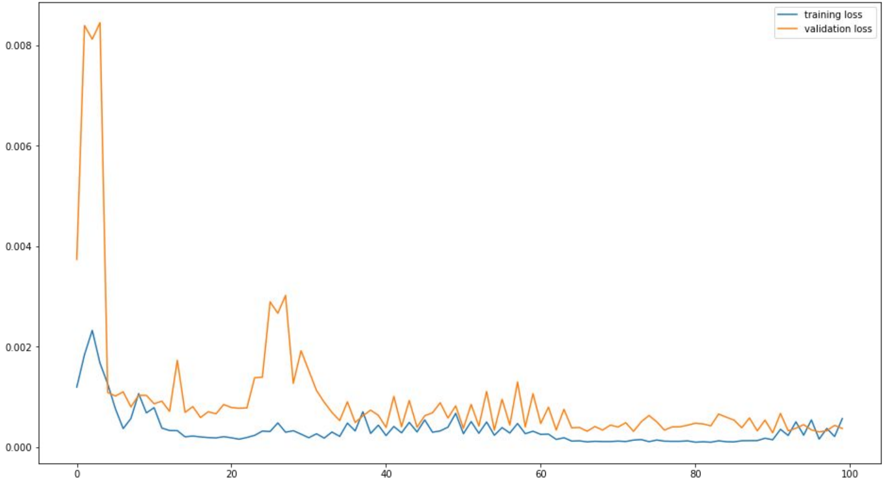
Loss Interaction

Epochs = 100 Batch Size = 32
Validation Split = 0.2 Verbose = 1

GRU



LSTM



Obtain Data

EDA

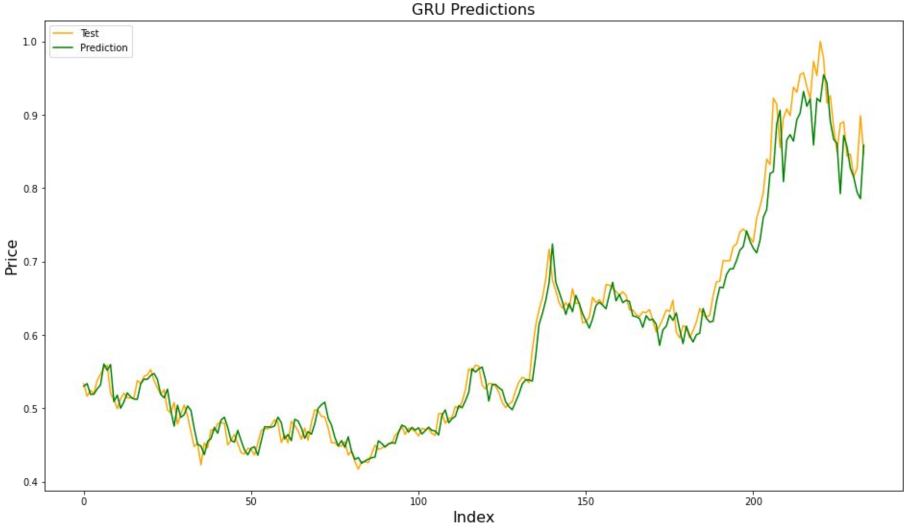
Pre-Processing

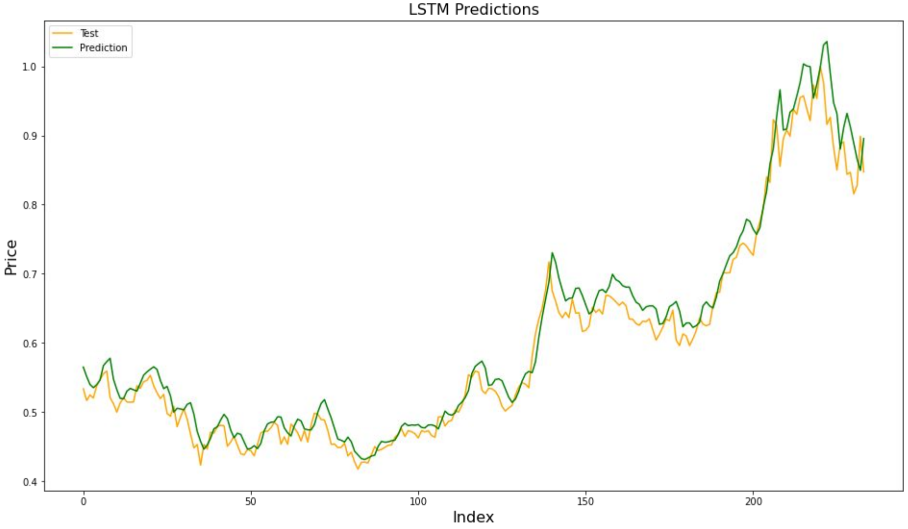
Modeling & Evaluation

Future Works

GRU and LSTM Model

80 - 20 Train Test Split





Obtain Data

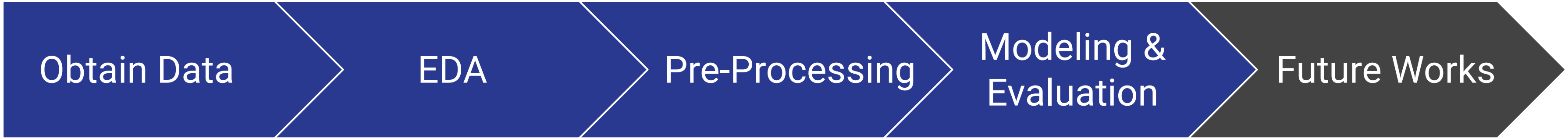
EDA

Pre-Processing

Modeling & Evaluation

Future Works

ARIMA	SARIMAX	GRU	LSTM
Metric Used : MAPE (Mean absolute percentage error)	**Metric Used : MAPE (Mean absolute percentage error)**	**Metric Used : MAPE (Mean absolute percentage error)**	**Metric Used : MAPE (Mean absolute percentage error)**
The average difference between the predicted value and the actual value is	The average difference between the predicted value and the actual value is	The average difference between the predicted value and the actual value is	The average difference between the predicted value and the actual value is
(5Y) 12.61%	(5Y) 13.88%	2.6%	3.53%
(2Y) 21.5%	**(2Y) 3.97%**		



Obtain Data

EDA

Pre-Processing

Modeling & Evaluation

Future Works

More Data

- Include News Sentiment data from
  - News Articles
  - Reddit posts
  - Tweets
  - Discord channels
- Add statistical weights on certain price movements based on specific chart indicators

Create a WebApp

- Provide data on any stock ticker that is search
  - Bullish/Bearish
  - Increase/Decrease (with %)
- Personal Screener
- Recommended stocks
  - Industry
  - High Volume
  - Momentum

AI Trading Bot

- Only when I can obtain a MAPE of less than 1% and at least a 70% win-rate can I then consider the bot to be viable.

Thank You