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

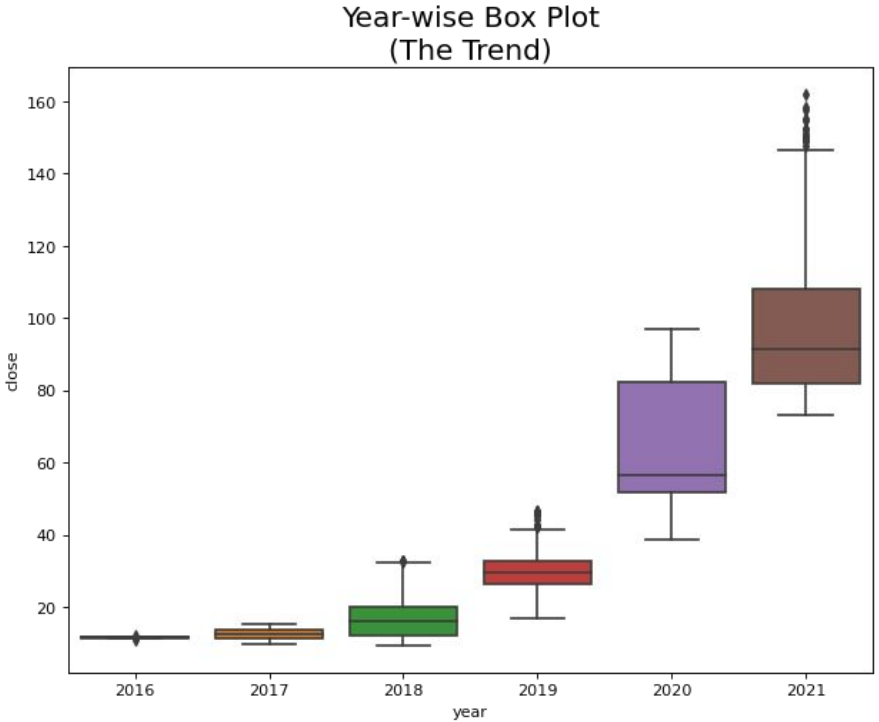

Obtain Data
EDA
Pre-Processing
Modeling & Evaluation
Future Works

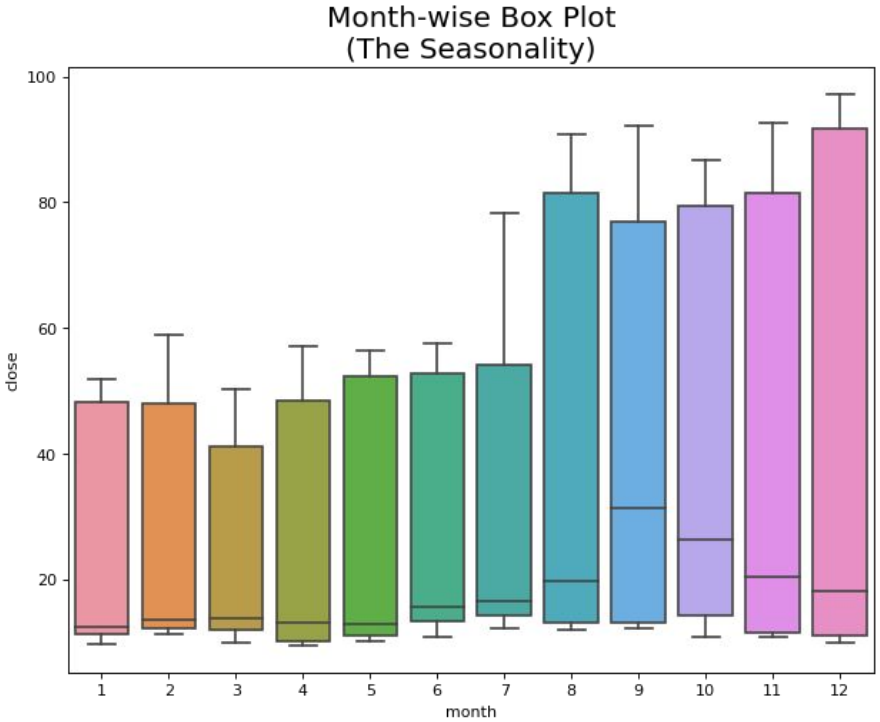
Pandas Datareader

- 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

Obtain Data | EDA | Pre-Processing | Modeling & Evaluation | Future Works





Adfuller Test #1

ADF Test Statistic : -0.1856205771364

p-value : 0.9402140424261141

Number of Observations Used : 1237

Comment : Time series is non-stationary

Apply Lag (24)

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

.

.

Adfuller Test #2

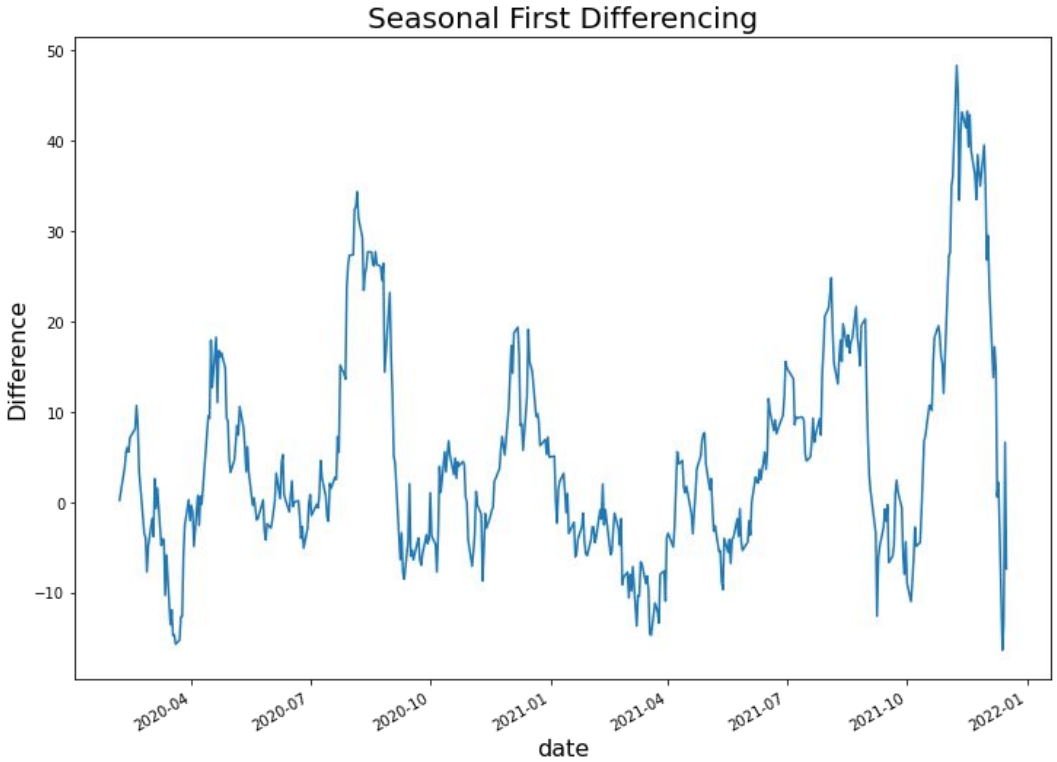
ADF Test Statistic : -5.439757731032

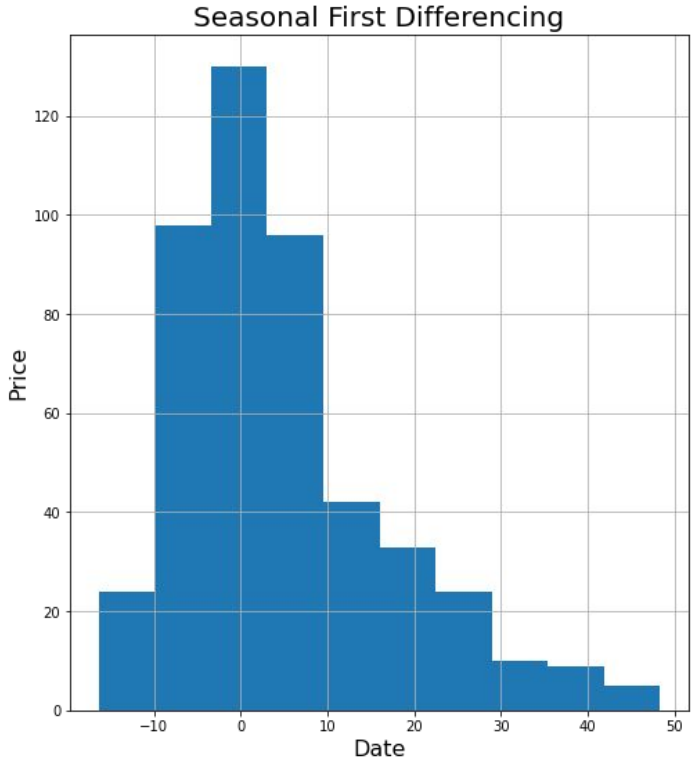
p-value : 2.793269722804705e-06

Number of Observations Used : 1209

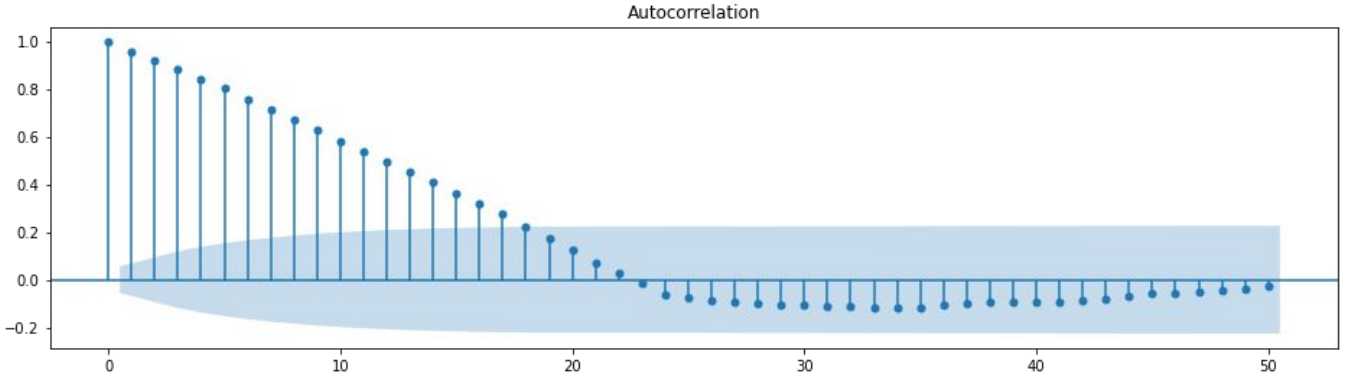
Comment : Time series is stationary

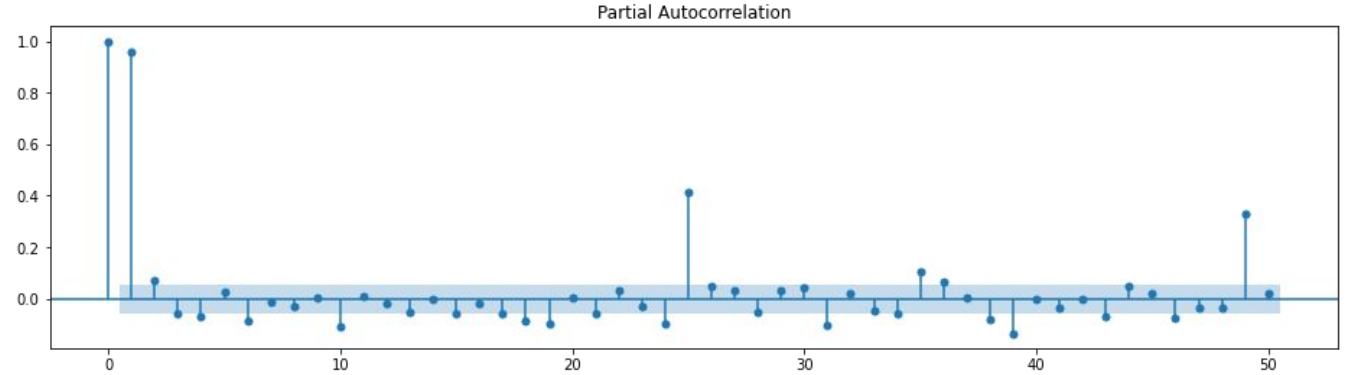
Obtain Data | EDA | Pre-Processing | Modeling & Evaluation | Future Works





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

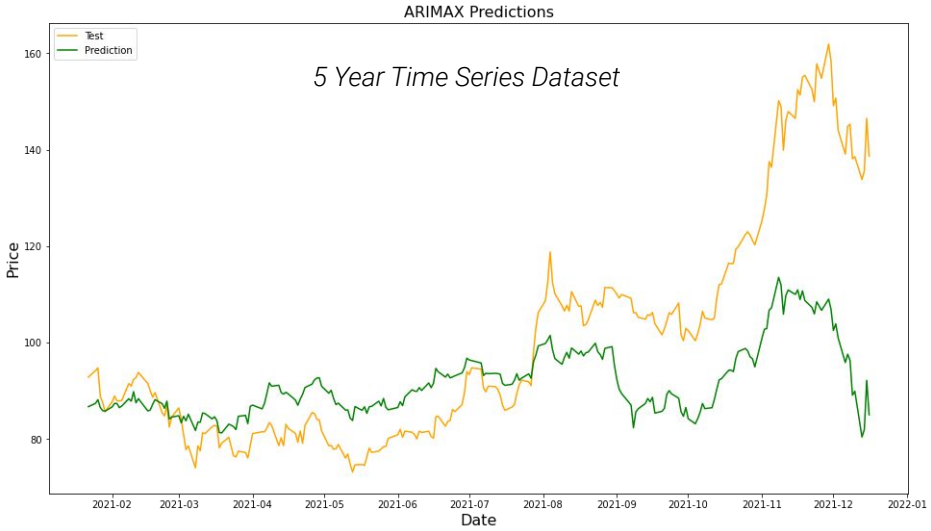
Evaluator

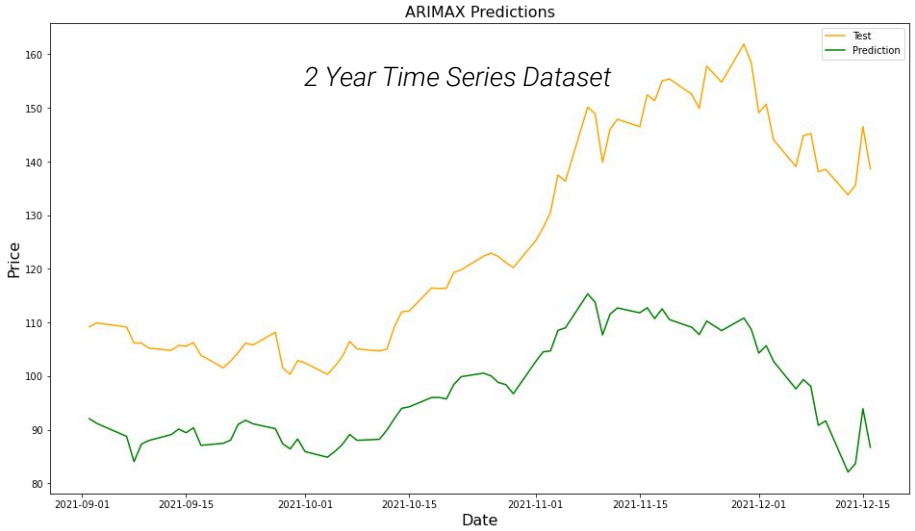
Loss : Mean Squared Error
Optimizer : Adam

Obtain Data → EDA → Pre-Processing → **Modeling & Evaluation** → Future Works

ARIMAX Model

80 - 20 Train Test Split

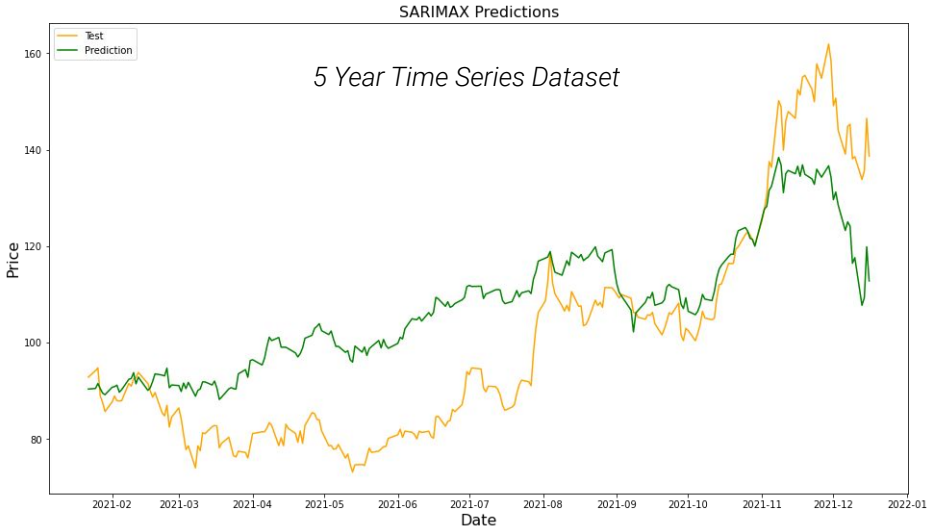


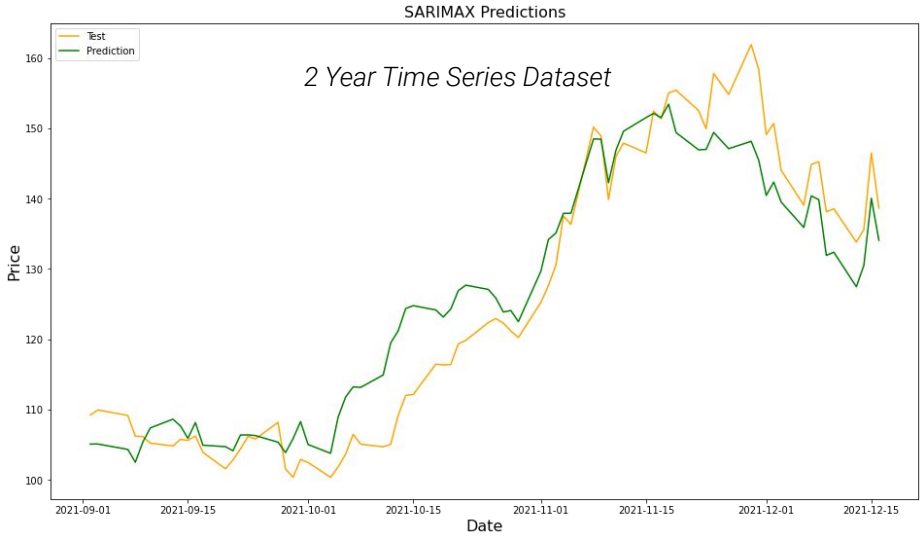


Obtain Data → EDA → Pre-Processing → **Modeling & Evaluation** → Future Works

SARIMAX Model

80 - 20 Train Test Split



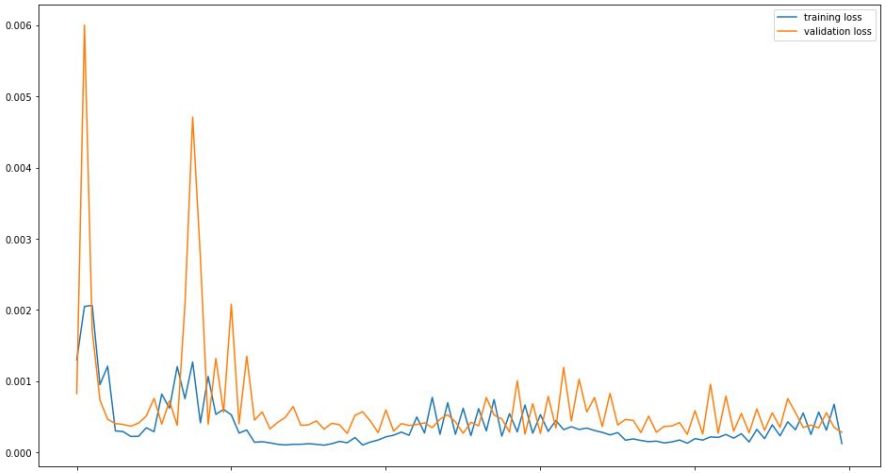


Obtain Data → EDA → Pre-Processing → **Modeling & Evaluation** → Future Works

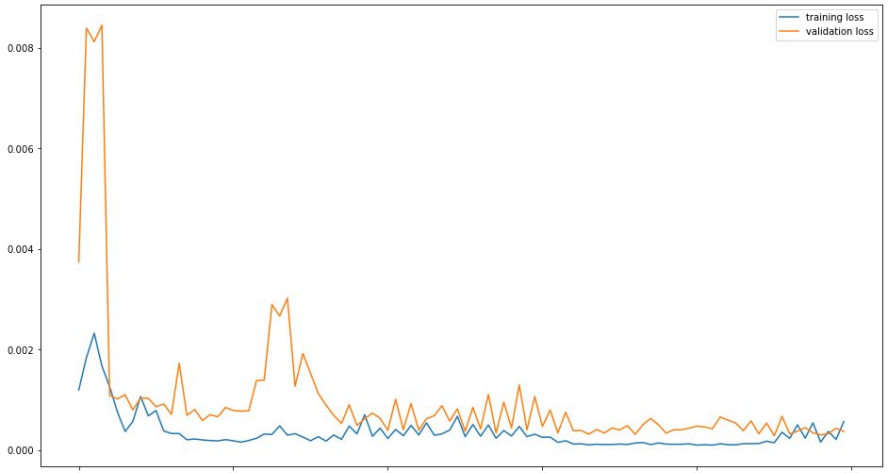
Loss Interaction

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

GRU



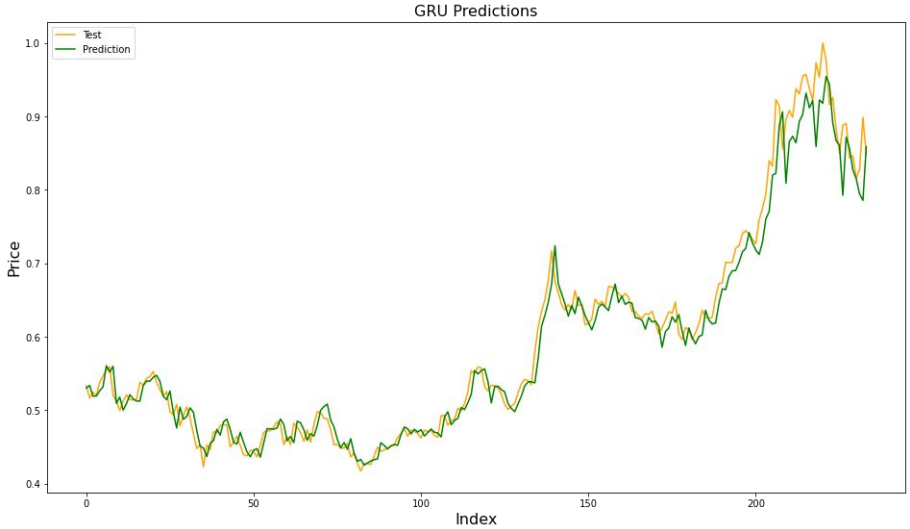
LSTM

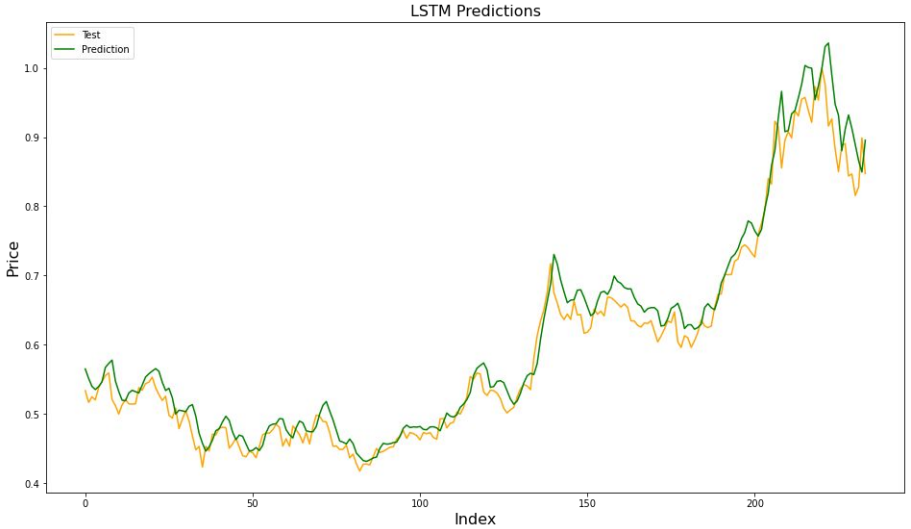


Obtain Data → EDA → Pre-Processing → Modeling & Evaluation → Future Works

GRU and LSTM Model

80 - 20 Train Test Split





ARIMA

Metric Used : MAPE (Mean absolute percentage error)

The average difference between the predicted value and the actual value is

(5Y) 12.61%

(2Y) 21.5%

SARIMAX

Metric Used : MAPE (Mean absolute percentage error)

The average difference between the predicted value and the actual value is

(5Y) 13.88%

(2Y) 3.97%

GRU

Metric Used : MAPE (Mean absolute percentage error)

The average difference between the predicted value and the actual value is

2.6%

LSTM

Metric Used : MAPE (Mean absolute percentage error)

The average difference between the predicted value and the actual value is

3.53%

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