### Experimenting with xgbclassifer to perform classification

did not work well, using apple price to predict apple movement only

Using only close and lag1 to 10

* Processing: 100%|██████████| 2262/2262 [21:23<00:00, 1.76it/s]
* CPU times: total: 2h 12min 23s
* Wall time: 21min 23s
* Accuracy: 17.20%
* A graph with red and blue lines

  Description automatically generated
* A blue squares with white text

  Description automatically generated

### XGBoost to perform regression

Using only close and lag1 to 10

Processing: 100%|██████████| 2262/2262 [02:44<00:00, 13.73it/s]

Wall time: 2min 44s

A lot more faster can the classifier: from 30 mins to 3 mins

Similar accuracy Accuracy: 16.22%

A blue squares with white text

Description automatically generated

Accuracy: 16.22%

precision recall f1-score support

Moderate Downtrend 0.12 0.13 0.13 232

Moderate Uptrend 0.16 0.14 0.15 348

Slight Downtrend 0.21 0.18 0.19 464

Slight Uptrend 0.20 0.20 0.20 456

Stable 0.16 0.13 0.14 310

Strong Downtrend 0.13 0.17 0.15 218

Strong Uptrend 0.12 0.16 0.14 234

accuracy 0.16 2262

macro avg 0.16 0.16 0.16 2262

weighted avg 0.17 0.16 0.16 2262

A graph with red and blue lines

Description automatically generated

#### For now, just focus on using apple data to predict apple return.

1. To optimize the processing and ensure a more streamlined approach, we have adjusted the model to run on only 10% of the original time series data.
2. Using close price only, 1 – 20 lag close price, the accuracy 20.59%
3. After the addition of volume data, we observed an accuracy improvement. The model's accuracy decreased from 20.59% to 19.12%.

Instead of predicting next day’s return or movement, I decided to have a broader timeframe: predicting the next week movement or price. The goal of the model is to estimate the price of a stock will obtain in the following week, and by doing that, we can use the model to discover trading signal.

Xgboost model with apple price only, predicting the return 5 days (a week) after, with 5 lag close in total + current close

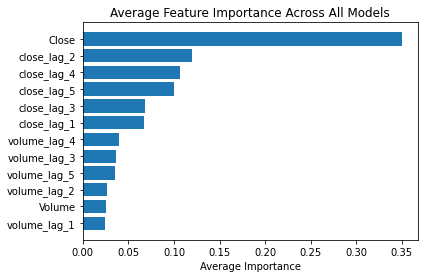
Mean Squared Error (MSE): 0.9973282042690228

Mean Absolute Percentage Error (MAPE): 2.84%

After adding volume and 5 lag volume

Mean Squared Error (MSE): 1.1600404210958313

Mean Absolute Percentage Error (MAPE): 3.12%



It seems like volume does not have much explanation power, which is counterintuitive

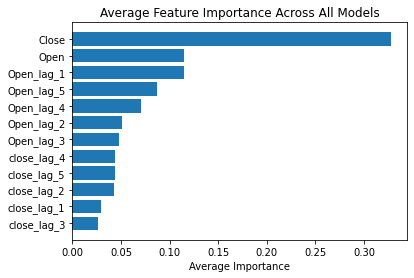
So combine volume and close price might be a good choice

After including open price and its lag price, the performance of the model slightly decreased

Mean Squared Error (MSE): 1.0266442298919725

Mean Absolute Percentage Error (MAPE): 2.93%

However: open seems also very important



Trying to add highs and lows of a day

Mean Squared Error (MSE): 1.063908647152127

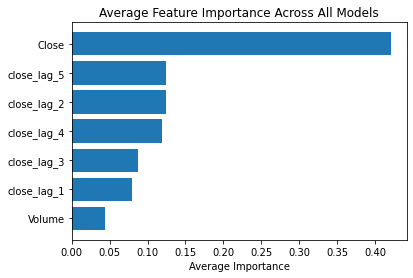
Mean Absolute Percentage Error (MAPE): 2.96%

Probability due to collinearity, so do not consider them at this moment

Added volume but not its lag; this is going to be our baseline

Mean Squared Error (MSE): 0.992164832481723

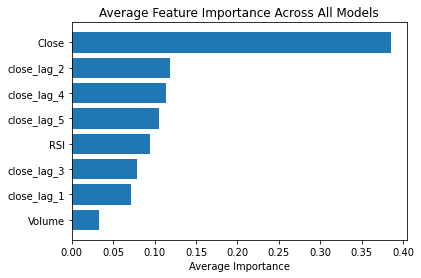
Mean Absolute Percentage Error (MAPE): 2.83%



After adding current RIS

Mean Squared Error (MSE): 0.9760073352410757

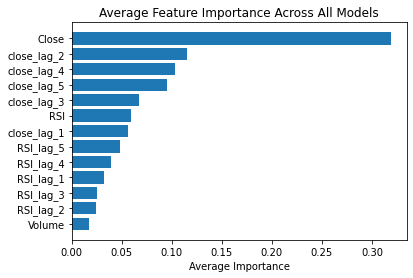
Mean Absolute Percentage Error (MAPE): 2.70%



Adding lag rsi

Mean Squared Error (MSE): 1.0700108377431683

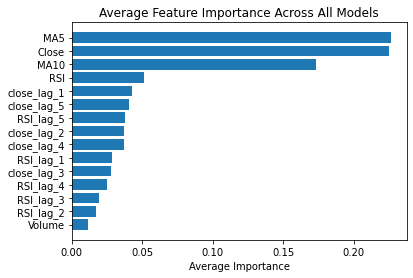
Mean Absolute Percentage Error (MAPE): 2.79%



Adding ma 5, 10

Mean Squared Error (MSE): 0.9630627719004787

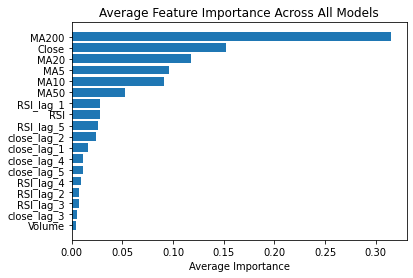
Mean Absolute Percentage Error (MAPE): 2.66%



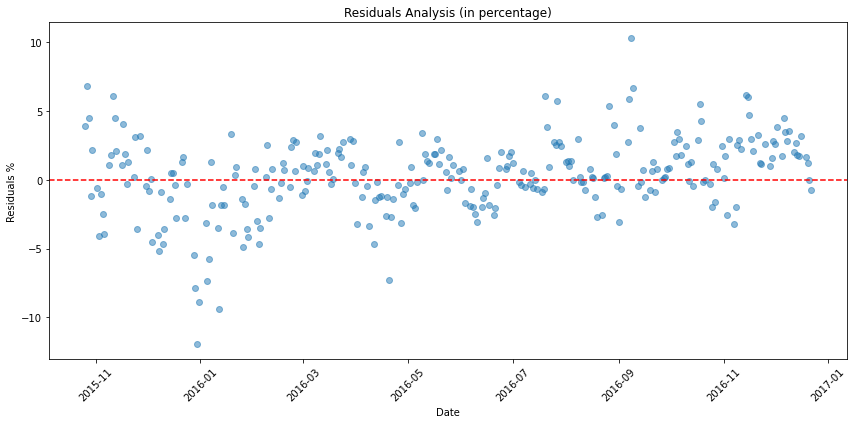
Adding all important mas, reference how to swing trade, huge improvement!!!!!!

Mean Squared Error (MSE): 0.5414648921605042

Mean Absolute Percentage Error (MAPE): 2.05%



The residual looks a lot more random than before



Adding a economic indicator, nearly no change, I expect this variables to be more powerful when there is a huge economic change (interest rate), so I am keeping it

Mean Squared Error (MSE): 0.5675471716347179

Mean Absolute Percentage Error (MAPE): 2.06%

Not using lag values for rsi

Mean Squared Error (MSE): 0.5752399599689242

Mean Absolute Percentage Error (MAPE): 2.02%

Adding a variable called green, small change

Mean Squared Error (MSE): 0.5728439463028074

Mean Absolute Percentage Error (MAPE): 2.01%

Add open close difference (open are high correlated with close)

Mean Squared Error (MSE): 0.5728439463028074

Mean Absolute Percentage Error (MAPE): 2.01%

Add high\_low\_diff \* greenday,

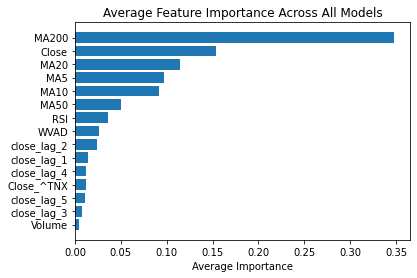
Mean Squared Error (MSE): 0.6062255531910087

Mean Absolute Percentage Error (MAPE): 2.05%

those can be covered using WVAD, removed those and added WVAD

Mean Squared Error (MSE): 0.5834007328317992

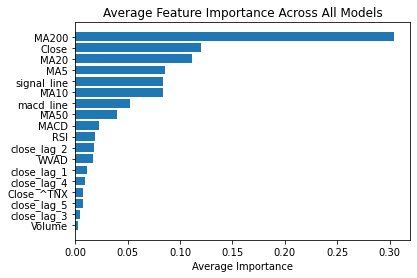
Mean Absolute Percentage Error (MAPE): 2.04%



Added macd, macd line and signal line

Mean Squared Error (MSE): 0.5025457907090828

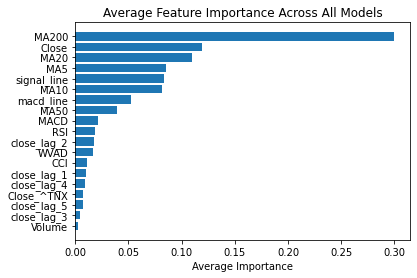
Mean Absolute Percentage Error (MAPE): 1.96%



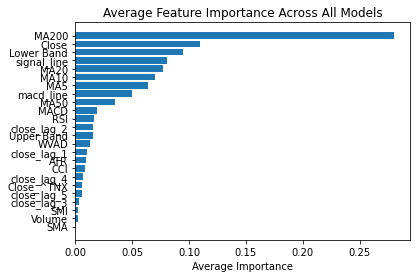
Added CCI

Mean Squared Error (MSE): 0.5008917319007433

Mean Absolute Percentage Error (MAPE): 1.95%



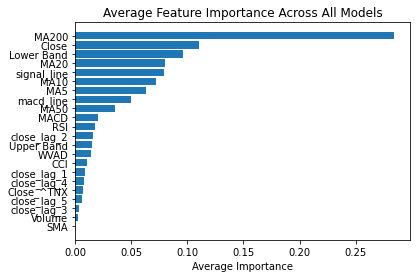
Added SMI and ATR, the model shows they are not important



Removed the previous two and Added BOLL

Mean Squared Error (MSE): 0.5331274985235674

Mean Absolute Percentage Error (MAPE): 1.93%

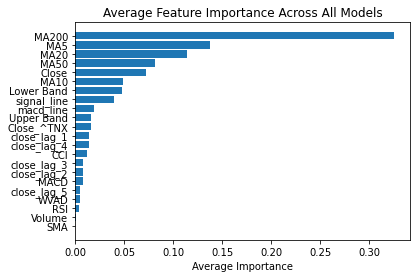


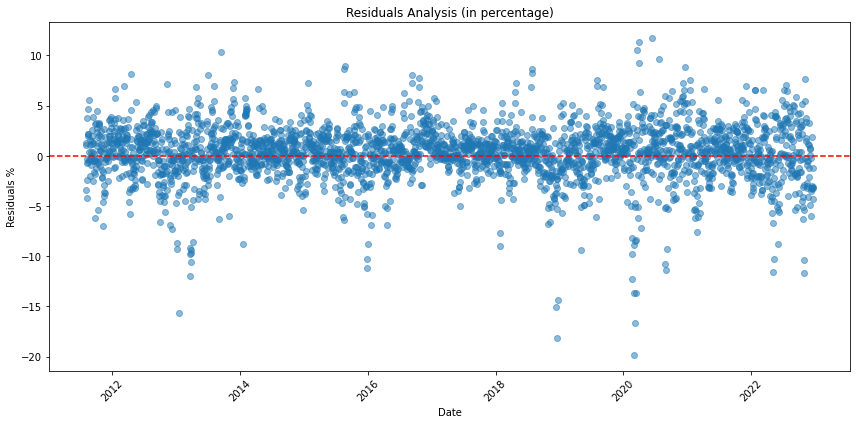
Take a break and see how to model runs on a broader time frame:

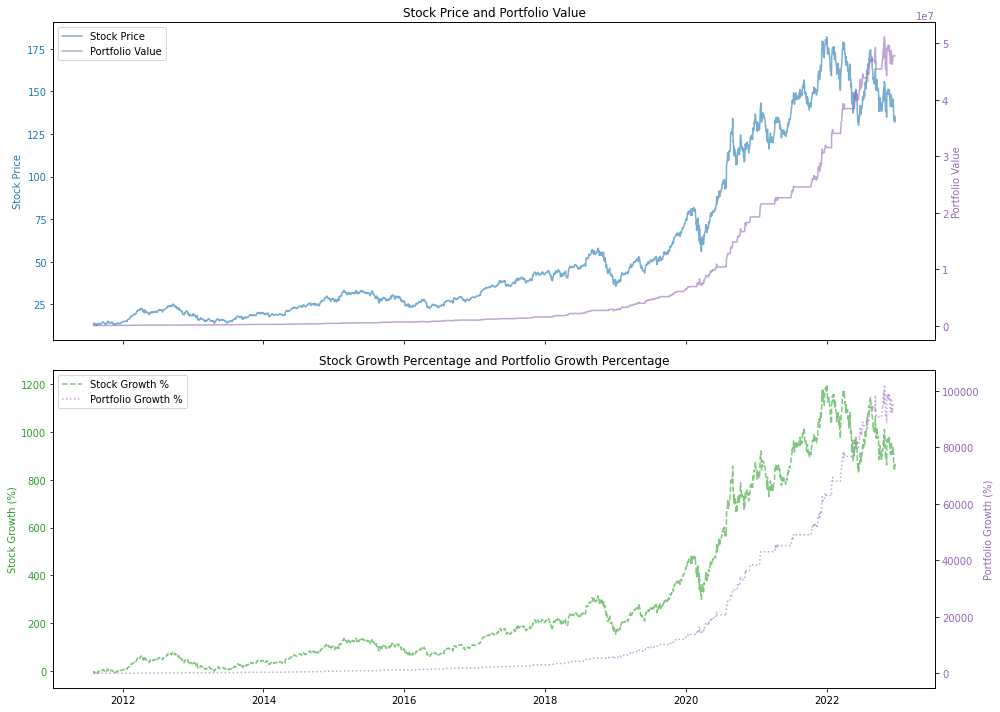
Processing: 100%|██████████| 2866/2866 [02:11<00:00, 21.72it/s]

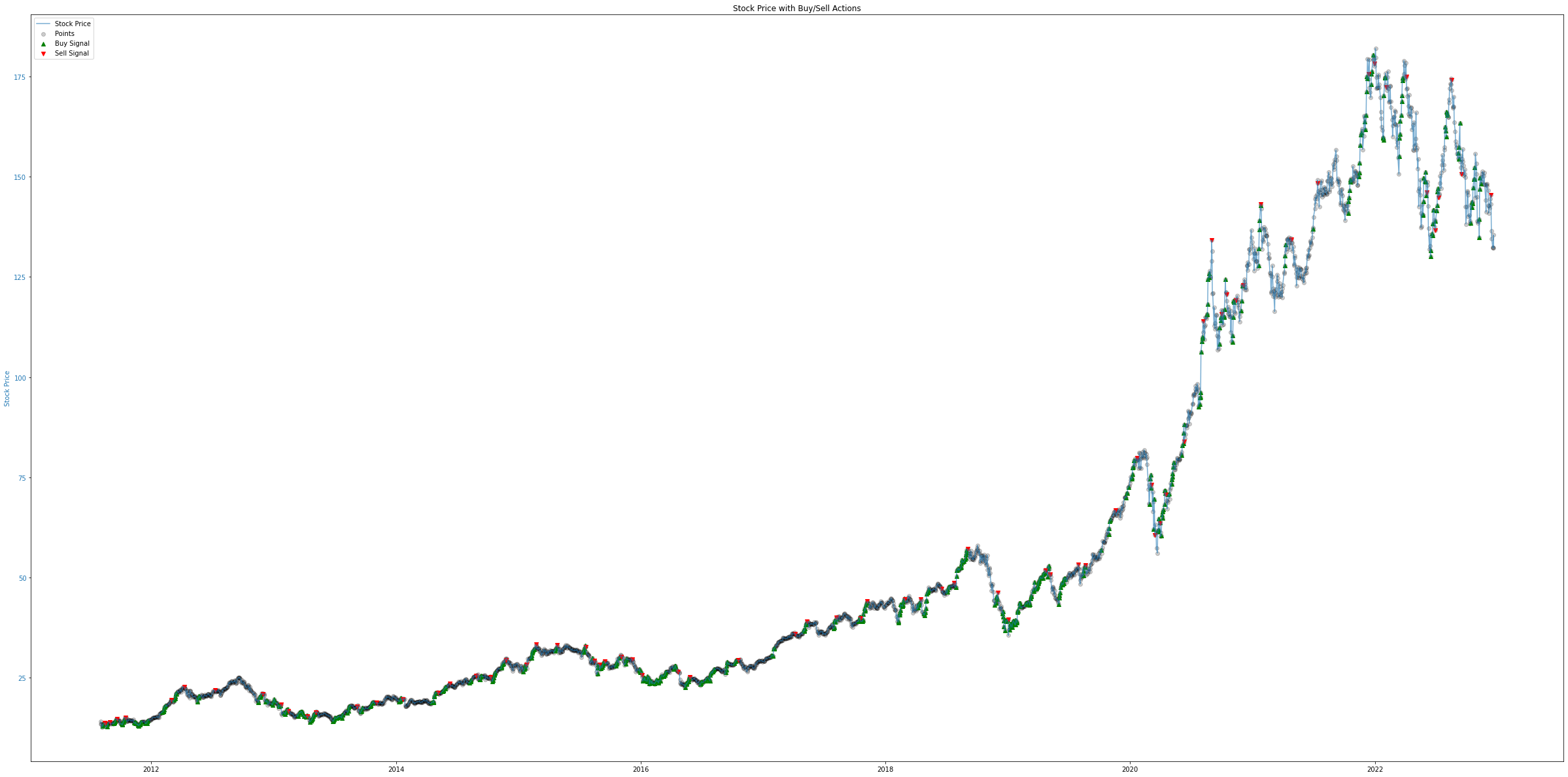
Mean Squared Error (MSE): 4.8594886373102595

Mean Absolute Percentage Error (MAPE): 1.95%









Decent at capture selling signal

Should also consider Williams vix fix and rsi bol strat