Justin Rebollo

CS 677

HW 2

**Question 1-**

1. See code in file: rebollo\_hw2\_1.py

SPY = pd.read\_csv("SPY.csv")

RTX = pd.read\_csv("RTX.csv")

SPY.loc[SPY['Return'] >= 0.0, 'True Label'] = '+'

SPY.loc[SPY['Return'] < 0.0, 'True Label'] = '-'

RTX.loc[RTX['Return'] >= 0.0, 'True Label'] = '+'

RTX.loc[RTX['Return'] < 0.0, 'True Label'] = ‘-‘

2. See code in file: See code in file: rebollo\_hw2\_1.py

True Label

+ 0.567117

* 0.432883

3 & 4. See code in file See code in file: rebollo\_hw2\_1.py

K Values for - days repeating

SPY: K = 1 60.37%

RTX: K = 1 54.16%

SPY: K = 2 59.26%

RTX: K = 2 54.44%

SPY: K = 3 59.09%

RTX: K = 3 59.35%

K Values for + days repeating

SPY: K = 1 53.86%

RTX: K = 1 52.32%

SPY: K = 2 46.48%

RTX: K = 2 47.28%

SPY: K = 3 51.71%

RTX: K = 3 49.46%

**Question 2-**

1. See code- rebollo\_hw2\_2-5.py

W Ticker TP FP TN FN accuracy TPR TNR

2 SPY 263 192 17 33 0.55 88.85 8.13

3 SPY 264 192 17 32 0.56 89.19 8.13

4 SPY 134 101 108 162 0.48 45.27 51.67

2 RTX 31 24 211 239 0.48 11.48 89.79

3 RTX 24 18 217 246 0.48 8.89 92.34

4 RTX 127 128 107 143 0.46 47.04 45.53

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

2. The accuracy is computed below

SPY: Year 2- 55%, Year 3-56%, Year 4- 48% RTX: Year 2- 48%, Year 3-48%, Year 4- 46%

3. W = 3 has the highest accuracy for SPY and W = 4 has the highest accuracy for RTX.

**Question 3-**

1. See code- rebollo\_hw2\_2-5.py

2. See below for the accuracy of the ensemble predictions:

Ticker TP FP TN FN accuracy TPR TNR

E SPY 192 136 73 104 0.52 64.86 34.93

E RTX 0 0 235 270 0.47 0.00 100.00

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Question 4-**

1-7. Please see the table below and the code in rebollo\_hw2\_2-5.py-

W Ticker TP FP TN FN accuracy TPR TNR

2 SPY 263 192 17 33 0.55 88.85 8.13

3 SPY 264 192 17 32 0.56 89.19 8.13

4 SPY 134 101 108 162 0.48 45.27 51.67

E SPY 192 136 73 104 0.52 64.86 34.93

2 RTX 31 24 211 239 0.48 11.48 89.79

3 RTX 24 18 217 246 0.48 8.89 92.34

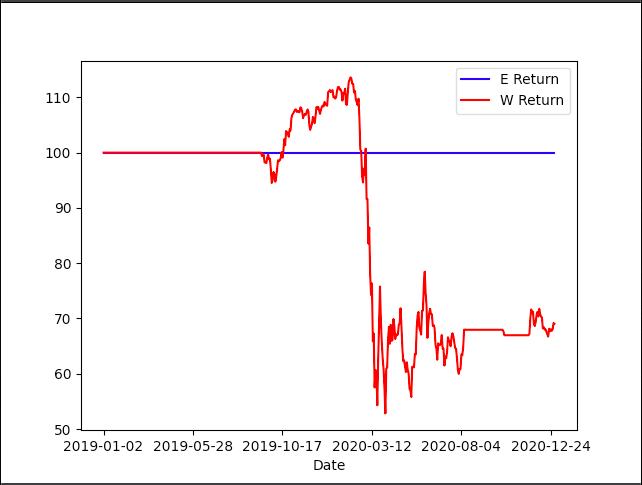
4 RTX 127 128 107 143 0.46 47.04 45.53

E RTX 0 0 235 270 0.47 0.00 100.00

8. Discuss your findings:

The W = n models seem to provide the same level of accuracy amongst SPY aside from an overall drop at W = 4. The opposite trend holds for RTX where the accuracy increases at W = 4. The ensemble method for SPY scores below the best preforming W = n model while the ensemble method is slightly better for RTX. It seems like the RTX Ensemble model is skewed and is creating a large TNR value. The ensemble method for SPY seems to be more representative of the data with satisfactory TPR and TNR values.

**Question 5-**

1.  Please see the below plot of the returns and see rebollo\_hw2\_2-5.py:
2. examine your chart. Any patterns? (e.g any differences in year 4 and year 5:

The graph above reflects the best W return which is W = 4 for RTX and its corresponding Ensemble return method. For the first year, the W = 4 method works great until about the end of the first quarter of 2020 where the model’s returns are significantly lower than before. In context this makes sense and might be a reflection on the economic downturn of 2020 which began in March, the same month of the return downtown, more so than a reflection on the trading strategy. The ensemble method seems to have a steady return because the selling of the stock and rebuying is not triggered when there are consecutive negative days in the predicted ‘True Label column’.