**Graphical user interface

Description automatically generatedHighly Correlated VariText, letter

Description automatically generatedables**

**Method for Removing Highly Correlated Variables:**

I created a dictionary where there was a key for every variable that was highly correlated (above a threshold of 0.97) with another variable. The value associated with each variable-name key in the dictionary was a set of the variables the key was highly correlated with. I then fitted a Random Forest on the data and found the feature importance. I iterated through the features in order of their importance and maintained a set of variables to drop. If I came across a variable that was in the drop list, I ignored it. If I came across a variable that was not in the drop list, but was in the dictionary, I extended the drop list with the variables associated with that key. In the end, I had a list of variables that I would no longer use in my analysis. That list is as follows:

**Graphical user interface, text, application, email

Description automatically generated**

**Hyperparameter Tuning of the Random Forest**

I used the library Hyperopt to tune the Random Forest Hyperparameters. I tuned the number of estimators, and the minimum number of samples required to split a leaf node. The optimal results yielded a mse of 0.01522 using 250 estimators and 14 samples. The search space was composed of 50, 100, …, 1000 estimators, and 2, 4, …, 20 samples. The final result has around 60 % accuracy of guessing whether a stock will go up or down.

**Shorting Stocks**

**Chart

Description automatically generated**I added shorting logiq to the weekly stock back tester. My first attempt is documented in the following graph. The Bot performs well at the beginning, so I was optimistic that if I included retraining, that the bot would continue to perform well. It seemed like the predictions might have deteriorated with time.

However, even when retraining the bot every 200 weeks, the performance still wasn’t satisfactory. It’s curious that the bot does so well in the beginning. It reflects some kind of potential.

**Chart

Description automatically generated**

**Chart, histogram

Description automatically generatedChart, scatter chart

Description automatically generated**

Figure : MSE over time.

Figure : Scatter of rate of change of my portfolio vs snp.

Bot Performance Over Various Testing Weeks

**Chart, line chart

Description automatically generated**

Clearly the more data the bot gets as input, the better it performs. Maybe we should just use the maximum number of weeks that is feasible for the dataset.