**Weekly Update**

**Nov 2- Nov 6**

Summary of work:

 Early in the week, we had a meeting with Cary about further assumptions and areas of confusion. An interesting merging issue occurred that delayed some progress this week, mixing up dates for some tickers and providing two rows with the same ticker and month. After meeting with Cary, we decided as a group about how to incorporate Cary’s expectations to our project and the majority of our work done this week involved deciding on the models to use, splitting up into groups to begin sub-tasks, and starting to build our respective classifier and regression models.

Notes from Cary Meeting:

We started our meeting by presenting Cary with the final data table that we created during the past two weeks and explained to him the meaning of some new X variables that we added since last meeting. Then we confirmed with Cary on what are expected in terms of the model and stock portfolio, making it clear about what his expectations are: he wanted to see around 30 to 50 companies with job metrics that are highly correlated to the stock data in the model, in other words, a selection of companies that can best represent the value of Greenwich’s data. In addition, he also mentioned that our trading strategy does not need to include every company, neither does it need to keep all companies throughout. Ultimately, Cary wants a dynamic trading strategy that can change the holding position of portfolio stocks from time to time based on our prediction models.

Additionally, Cary highlighted the importance of salary data usage in the model and requested that we should take careful consideration before modifying this job metric in that it is a metric that their competitors do not provide. We will definitely aim to showcase this comparative advantage of Greenwich when building our model.

Lastly, Cary suggested us to include lagging time effects in our model and try to find a way to treat companies with lagging effects differently than others. Cary suggested us to try in the range of 1 month lagging to 5 months lagging and we agreed upon a plan to try different combinations of X variables and stock returns from different lagging time span, ranging from 1 month to 5 months.

Plans for next week:

Based on the meeting notes, we also organized an internal team meeting to split up the work to do. Currently our plan is to break up the task into two smaller groups. Matt and Isaac are in charge of the SVM model, while Binqi and Congda are responsible for the regression model. For each model, we would first do a train-test split (After our meeting with Cary, we decided to initially pick the final 3 months of our data as the test set.), then use k-fold cross validation to try different combinations of hyperparameters to find the optimal model.

Also, we would apply our models to the dataset based on different lagging effect assumptions and figure out the most appropriate lagging time span. After we get the training result, we would conduct some statistical analysis on the result, and try to choose the appropriate benchmark of predicted stock return for picking stocks for the portfolio.