sddec18-13: Asset management - Financial Factor Discovery - "Value"
Week 9 Report
March 29 - April 4

Team Members

Carter Scheve — *Communications Lead*
Nathan Hanson — *Project Progress Tracker/Manager*
Caleb Utesch — *Meeting Scribe*
Jack Murphy — *Research Analyst*
Samuel Howard — *Lead Engineer*
Alex Mortimer — *Project Manager*

Weekly Summary

This is the ninth weekly report for our senior design project, and we have begun preparations for the end of the year tasks, including making final revisions to our design document and preparing for our presentations to both the class panel and our client team. We have started to shift away from exploratory tactics, moving more toward describing the progress made in model development and feature selection. We need to have concrete evidence of the work that has been done and where we plan to go next semester, so we will be collecting our results in all areas during the next couple of weeks and preparing them for presentations.

Summary of Progress this Report

- Integration of PCA into data importing library
- Provide model predictions with rolling window and expanding window training techniques
- Begin preparing end-of-semester presentations for class and for client

Pending Issues

- Difficult to balance further model exploration and end-of-semester preparation
- Model accuracies are not correct for some of the data

Plans for Upcoming Reporting Period

- Prepare summary of current models and feature selection techniques, both for class and client
- Prepare list of questions and concerns for on-site client meeting Tuesday
- Ensure all end-of-semester deadlines are known and on schedule

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours

Carter Scheve	For the data library, I worked on adding more functionality and fixing the README. Also doing testing and bug fixing for the data library. The train/test/split functionality also was modified to make the train and test sizes more controllable. I ran experiments for sliding and expanding window sets. I updated the README with these tests as examples for train/test split	8	88
Nathan Hanson	Continued work with classification confidence/probability scores. Began exploration of usage. Researched ways to combine classification and regression techniques for more complex decision information.	6	77
Caleb Utesch	This week I started working with univariate feature selection. Used sci-kit learn's built in univariate selection algorithm and ran it using a chi-squared statistical test. I chose to narrow down the features to the 10 most statistically significant, as our client suggested trying to narrow down the number of features to the range of 10-15. Once I ran this trial I compared the results with the features that were outputted from the tree-based feature selection that I'd been working with for the past couple of weeks	9	75
Jack Murphy	Continued to work with Recursive Feature Elimination. Did tests for prediction of Value and Momentum factors. I am beginning to collect an array of common features that show up during these tests. Each test yields fairly different results, but there are a few features that seem to show up each time. This week I was able to get 10 features that seem to be the most relevant for predicting the Momentum factor.	10	79
Samuel Howard	Since PCA showed such promise, I added the functionality to the data library. The function takes in the test and training data, then returns a transformed version with the desired number of components, either directly supplied or through the desired amount of variance explained.	10	80
Alex Mortimer	Used the data importing library to continue testing with random forest predictions, both	8	82

	in classification and regression models. Now that we have the data set trimming NaN values, instead of replacing them with mean or mode, the data is more representative of actual dataset, and is producing better results. This coming week I will be focusing on producing quality graphs to present to the client in our meeting, as well as ensuring my models are producing legitimate results.		