sddec18-13: Asset management - Financial Factor Discovery - "Value"
Week 10 Report
April 5 - April 17

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 tenth and final weekly report for our financial factor analysis project, spanning two total weeks. We have almost completely ceased work on the model tuning and feature selection that have been the focus for the semester, and moved our efforts to preparing the final presentations and reports for the class and client. This involves a lot of work in preparing graphs and other visuals to show the predictions and results of our models, as well as the results of the feature selection work done by the team. We have lots of work to put on display, but need to format it all in a way that is informative to the review committee and others with no knowledge of our project.

Summary of Progress this Report
- Met with client's professionals on site to demonstrate progress and get feedback
- Worked on revisions for final versions of the Project Plan and Design Document
- Finished weekly reports
- Began examining new data set and incorporating it into our data library system
- Worked on cleaning up current results for final presentations

Pending Issues
- The new data set has more than 5 times the number of columns than our original set, so it is difficult to visualize
- We need to normalize and organize the data into usable information for the models, including:
    - Discretizing some columns for classification
    - Using Principal Component Analysis to produce better columns
- Some of our project is confidential, and we are still figuring out what is allowed to be used in the final presentations. If the client wants us to leave some of our information out, we may have a hard time presenting information and answering questions from the panel
- We are being pressed to produce more results for the client, when the documents needed for class are taking up lots of time as well

Plans for Upcoming Reporting Period
- Analyze the new data set and make changes to fit requirements for the inputs to our models
- Prepare final presentations for the class and the client
- Revise project plan and design document for final submissions

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Carter Scheve	Worked with weighting model training data with different curves and graphing results of the accuracy difference Recreated windowed tests. This includes expanding and sliding window tests. I did these with varying parameters to get a full range of accuracy data over time Preliminary research into new dataset including exploratory research and comparison to previous dataset	14	95
Nathan Hanson	Began experimentation with different implementations of classification probability data. Learned usage of PCA data analysis techniques, as well as sliding and expanding window techniques on the dataset. Cleaned code repository, updated group's website. Made updates to project documents. Early research on methods to tune both classification and regression models, other than feature selection.	10	87
Caleb Utesch	Started looking into different possible statistical tests to use in univariate feature selection other than the Chi^2 test. Began looking at new dataset provided by client, and tried to figure out which feature selection methods might be good to run on this data. Will most likely conduct a similar process of running several different feature selection methods and comparing the results from each. This will hopefully lead to a comprehensive, useful set of features to input into our models.	11	86
Jack Murphy	Continued to work with the Recursive Feature Elimination method of feature selection. Looked into the new features provided by the new dataset our client has given us. Gathered an array of 10 features with a rank of 1. There are many different ways to tune this model and I have been collecting all of the various features that have been output. The new dataset has provided new features that will	11	90

	need to be analyzed.		
Samuel Howard	Investigated which feature accounted for the most variance, and their correlation. After continuing to see unusual results, I have decided to normalize the data to see if that might produce more expected data. The data after normalization met more closely with expectations. This means fifteen components are needed to explain 99% of the variance, a much more reasonable prospect than just one.	13	93
Alex Mortimer	I was out of town for most of the second half of this time period, but for the first half I worked primarily on cleaning up and refining my Random Forest models. Since I have been working more with the classification model the last few weeks, I have more consistent and sensical results from that. My methods of displaying the results (confusion matrices, accuracy scores, etc.) are underdeveloped, so I spent some time working on improving the way the information is shown. For the regressive model, I am still trying to figure out exactly what my model is using to make predictions, and what the error rate and accuracy scores mean in this sense. In the meantime, I am always using our data library and providing feedback for bugs and other ways in which it can be improved moving forward.	15	97