sddec18-13: Asset management - Financial Factor Discovery - "Value"
Week 8 Report
March 22 - March 28

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 eighth weekly report for our team, and we are busy working toward a full range of information to display at our end-of-semester presentation. During our client meeting last week, the client asked for us to provide lots of documentation about what progress we have made, centered around statistics for the accuracy of models. In addition to that, we were asked to do some more in-depth research into scholarly articles to give additional guidance about potential approaches to the optimization of our models, along with reports about these papers. This has kept us plenty busy but has given good insight about where we can push forward in the remaining weeks.

Summary of Progress this Report

- Further testing on Train-Test split function and custom data import library
- Read articles/papers on our own models and their use in stock predictions
- Reported our findings from the readings to the client for future reference
- Began changing our models to provide regressive predictions as well as classifying

Pending Issues

- Difficult to choose whether to pursue regressive or classifying models
    - Both have advantages and disadvantages
    - Both are used for different types of predictions
- Client is requesting ever increasing amounts of deliverables each week
- Request for research, development, and reporting needs to be prioritized per week

Plans for Upcoming Reporting Period

- Fix errors in our model prediction strategy as suggested by our client
- Produce more subsets of features that produce better results in each model
- Continue translating models from classification to regression

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. These functions included discretizing columns into classifications rather than discrete values. I also worked on sliding and expanding window testing for Bayes Classifiers.	11	81
Nathan Hanson	This week I did some experiments with some features chosen via Caleb's feature selection work. I also did some research, including a literature review, on the role of probability scores in classification models.	9	71
Caleb Utesch	Continued working on tree-based feature selection during this week, as recommended by our client. Ran several different trials putting in different parameters to the selection algorithm and compared the results from each trial in order to try and determine which, if any, features were appearing the most and had the most significance to our models	9	66
Jack Murphy	This week I looked into recursive feature elimination as another method of feature selection. I was able to get a ranking of the features in our data set in order to see which features could potentially be best to use for our models.	10	69
Samuel Howard	This week I looked into principal component analysis as a method of feature selection. It surprisingly effective, given one component can be used to explain over 99% of variance within the dataset. This means we can use one dimension to account for over 99% of information contained within the dataset. Further testing will be needed to ensure no drop in quality is observed.	10	70
Alex Mortimer	Continued reviewing the data importing library and providing feedback. Read a paper about developing a Random Forest model to predict stock performances and gained some insight about possible techniques to try, as well as an in-depth understanding of how the Random Forest method works. Developed a new regressive Random Forest	12	82

	model, similar to the classification model, and began examining results and ways to optimize.		