EE/CprE/SE 492 Biweekly Report 4
October 9 - October 22
sddec18-13: Asset management - Financial Factor Discovery - "Value"
Client: Principal Global Investors
Advisor: Chinmay Hegde

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*

Biweekly Summary of Progress

This is the fourth bi-weekly report for our senior design class. Since our last report, in which the the data aggregation portion was completed, the model application to the data has been started. We are taking the models we explored last semester and adjusting them to fit the new data with mixed results. Feature selection has also been applied to the new data as well, with differing results.

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Carter Scheve	Initial feature selection research, model research, midpoint presentation, volatility calculation, naive bayes research Started and completed data aggregation for volatility. Saved the data frames into a shared folder for reuse. Did initial research into other potential models and feature selection techniques for later use. Completed parts rough draft and final draft for client midterm presentation.	17	53
Nathan Hanson	Created a python module to create, serialize, read, and write copies of	15	47

	dataframes with data aggregated to factor level data. Began research on advantages of XGBoost, as well as potential alternative machine learning data targets. Began analysis of factor level data.		
Caleb Utesch	Continued implementing Tree-based feature selection with new dataset. Got it to work when filling all NaN values with 0s, but results don't seem to be very useful. Still working through how to replace NaN values with median values for that specific feature, and get subset of features from that.	14	46
Jack Murphy	Worked on Recursive Feature Elimination to start to compile a subset of features to work with. Will be able to do further analysis once our models are developed. Working on tuning the RFE model and analyzed how the different feature subsets affect overall prediction results.	15	50
Samuel Howard	Worked on PCA to further analyze the dataset. There were issues with applying the process to the original decile features. Additionally, work has started on a tool for use on a train, test split. As always, further testing is required to fully ascertain the usefulness. Additionally, autoregressive models were briefly tested, but verified to be as useless in this situation as the last one.	16	50
Alex Mortimer	Spent lots of time preparing the midpoint presentation for our on-site visit to our clients this Wednesday. Took suggestions from clients and worked toward a more accurate Random Forest Regression Model. Having issues with creating a valid model with realistic predictions, need some time to debug and work with the client to improve it.	19.5	58

Pending Issues
- Model accuracy continues to be an issue. At least one model reports $|r^2| > 100$.
- Definitions of certain predictors are difficult to translate into code
- Large portions of the data are thrown away at client's direction. This will likely reduce accuracy of models later in development.

Plans for Upcoming Reporting Period
- Obtain a final DataFrame of feature-level data to use with models
- Continue tweaking our models to better fit and use the data to improve results
- Continue making presentations for the client, including a mid-point presentation that will be on site in Des Moines
- Improve feature-analysis techniques with new data
- Figure out how to better store the aggregated data than using the provided server so that simultaneous work can be done