EE/CprE/SE 492 Biweekly Report 3
September 25 - October 8
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 third bi-weekly report for our senior design class. Since our last report, in which we spent lots of time clarifying expectations for deliverables with our client, we have had a more concrete progress level. Our team worked on all aspects of the predictive analytic process, including data aggregation from stock level to feature level, feature analysis to identify the most useful statistics, introductory runs through the predictive models to get out-of-box results, and descriptive analysis to present our findings to the client. As always, we have lots of paperwork to do for the client which takes up time, but we have gotten a lot more done in terms of developing code within the last two weeks.

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Carter Scheve	Contributed some work to the second part of the aggregation step in our EDA. Worked on setting up a data structure to use for the extra statistical calculations more efficiently. Started and made a lot of progress on a new aggregation for our data in 1, 3, 6, and 12 month volatility. Set up a meeting with our advisor and met with him as a group. Communicated with clients about challenges, questions and clarifications for the tasks	15	36

	assigned		
Nathan Hanson	Worked on query development for more efficient data acquisition; investigated an issue causing unresponsive server issues on the EC2 instance. Revisited some models from last semester's work. Investigated methods for caching database results. Began implementation and testing of handling high memory usage on EC2 instance.	12	32
Caleb Utesch	Worked with the rest of the team on formulating queries to create the necessary data frames required for input into our predictive models. Reviewed feature selection techniques worked on last semester. Such as Tree-based selection, Univariate selection, and L1-based selection. Started planning implementation of these techniques on the newly formed data frames that we've been developing over the last few weeks.	12	32
Jack Murphy	Looked into aggregating the data into Data Frames while using all of the features from the data set. Worked with the team to develop the necessary query for the database. Identified multiple columns that have Nan or none values. Started looking back into work from last semester on RFE for feature selection and feature elimination. Will begin to create a subset of data features to run through our models	14	35
Samuel Howard	Formalized the process to aggregate the data and create decile level features. Additionally, work has resumed in applying PCA to the data from last semester. Preliminary results suggest that the data is about as correlated as the last dataset, allowing reduction down to 5 axi for 99.9% variance.	14	34
Alex Mortimer	Took an in-depth look into the process of	18	38.5

	aggregating the stock-level data into feature-level data and creating DataFrames from that data to use in our predictive models. Developed an algorithm to aggregate the full stock-level datasets from multiple database tables into a DataFrame that has a combination of weekly data for each row. Calculated the mean, median, mode, standard deviation, and volatility of each feature in the weekly row and added those columns to the DataFrame. Used the newly calculated values to run through our original models and see preliminary results. Created various presentations to present these findings to our client.		

Pending Issues
- The client is set on creating a feature from the data that involves querying several weeks' data for each individual week, but the algorithm they provided has some issues which we will have to work through with them
- We are still confused with the purpose of the work we are doing, since it has already been done by the client. It seems redundant besides the pure educational value, but it is what the client requested so we're going with it
- The resource provided by the client struggles to work when the whole team is working at the same time. It's not an issue that we have much control over, but worth reporting since it hinders our ability to work simultaneously

Plans for Upcoming Reporting Period
- Finish data aggregation methods
- 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