sddec18-13: Asset management - Financial Factor Discovery - "Value"

Week 5 Report
February 22 - February 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 fifth weekly report for our team. This week continued to be heavy on Exploratory Data Analysis on the dataset provided by the client. We have been working hard to try and exploit trends between financial factors, but it has proven difficult since the signal between factors is very low. Obviously that is why predicting the stock market is so difficult, but we have been able to find some small trends amongst the data. Usually we present these findings to our client, and they tell us that either the findings are invalid or that they're obvious in nature, but the act of exploring and looking for trends has helped us understand the dataset better.

Summary of Progress this Report

- Continued carrying out EDA with each of our specific prediction models
    - Created report for client presenting each group members findings from individual EDA
    - Analyzed results from EDA and tried to make sense of what each individual group members findings meant
- Pitched EDA results to client
    - Went over advantages and disadvantages of each model as they applied to the dataset provided
    - Speculated on some experimental data analysis findings and what they could mean
    - Our client informed us of an error we were having when splitting input data into training and testing sets that involved needing to leave a one month buffer between the end of the training set and beginning of the testing set

---

Pending Issues

- Design document is due soon, and takes away time for the development we need to be doing
- All of our models need to implement a time buffer for the train-test split to validate results
- We may need to move our k-nearest neighbors and Support Vector Machines to solely feature selection purposes

---

Plans for Upcoming Reporting Period

- Revisit our models and continue trying to fit them to the datasets provided
- Implement the time buffer for train-test splitting
- Continue trying to validate the accuracy results of our models
- Finish the design document

- Create another report about our findings for our client to present in weekly meeting

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Carter Scheve	Learned more EDA on Kaggle.com. Including advanced filtering of data into groups Did more EDA on dataset given to us. Used more advanced techniques to separate data. Did the some of the same experiments on the correct data to be used in the model Redid basic machine learning experiments on correct data Did more advanced experiments with machine learning on the dataset. Used different sizes of permutations of the dataset and certain types of data to see what features might be selected in the future. Did more testing on data library to make sure functionalities were still working	11	50
Nathan Hanson	Created python module for reuse of SVM learning for extensive experimentation. Began naive experimentation with parts of data set. Began planning for automated testing of permutations of dataset. Explored options for visual demonstration of model performance.	13	46
Caleb Utesch	Took notes during all meetings and shared with the team on google drive. Continued conducting EDA using linear regression models. Ran several different trials to generate different models with inputs from all 5 different parameters categories to try and begin identifying any basic correlations. Will start looking into variable selection methods to try and identify what parameters might be best to use in future models.	9	41
Jack Murphy	Did more research into K-Nearest Neighbors model. Worked on manipulating how the test_train_split is performed. Did more research on general EDA with our dataset. Began to implement the data importing library to use with my experiments.	10	44

Samuel Howard	Researched more on autoregressive models. Started trying to apply them and related data analysis to the provided dataset. Found out there is some autocorrelation, will try to find maxima for positive and minima for negative correlation to try and tune the model. Will take a bit due to training time.	10	44
Alex Mortimer	After finding out some basic issues with how I had implemented the Random Forest model, I worked mostly on correcting those issues, and performing more exploratory data analysis. I found some trends while doing the EDA, but since we don't have much information on what the investments mean, I'm not sure if the findings are useful or not. That is something we discussed in our on-site meeting Wednesday, so we have a bit more clarity on that now.Moving forward, I just have to continue improving my model implementation and ensure it is correct before beginning to optimize.	11	52