sdmay19-06: Quantitative Research Modeling Library
Week 2 Report
September 15 - September 21

Team Members
Josiah Anderson — *Meeting Facilitator*
Doh Yun Kim — *Scribe*
Gabriel Klein — *Report Manager*
Drake Mossman — *Communication Manager*
Nathan Schaffer — *Overseer*
Jacob Richards — *Quality Assurance Manager*

Summary of Progress this Report
This week our team focused on examining existing scripts provided by Principal to better understand their processes. This included re-familiarizing ourselves with Python and looking into several libraries such as numpy, pandas, and sklearn to understand their usage. We also continued to work on getting interviews set up sooner than later so that we can narrow the scope and begin coding the project.

Past Week Accomplishments
Josiah Anderson
- Developed and documented project's problem and concrete deliverables
- Researched speed and efficiency differences between R and Python for large datasets
- Created presentation for client meeting

Doh Yun Kim
- Developed a risk management chart for the project
- Examined Principal's scripts to understand the processes
   - Looked into sklearn functionality
   - Tried to understand RBF, RationalQuadratic, and Matern guassian_process functions
- Continued learning Python
   - Parsing data from .csv files

Gabriel Klein
- Brought Python abilities back up to a satisfactory level
   - Got familiar with common functions to manipulate data such as zip and map
   - Reviewed built-in data structures and their features
- Studied scripts provided by Principal
   - Looked into numpy and ndarrays
   - Looked into pandas and dataframes
   - Began investigating sklearn and its various models and functionality
      - KNeighborsClassifier
      - RandomForestClassifier
   - Became familiar with overall flow and intent of the rolling window script

Drake Mossman
- Got familiar with Pandas Dataframes and their usage in the rolling window script

- Read basics of scikit learn library and relevant machine learning concepts used in example code
- Documented questions about the test scripts
- Investigated other example code, data, and output as received from the client

Jacob Richards
- Learned basic Python
- Went through Python example scripts from Principal
- Began investigating R

Nathan Schaffer:
- Researched and experimented with Python
- Looked over example code provided by Principal
- Researched, setup, and presented project management tools for the group to use
    - Primarily using GitLab's issue tracking and boards system

Group:
- Planned interview sessions with Principal for the first week of October

Pending Issues
In order to further refine the scope of the project, we still need to interview several more employees from Principal. We have these interviews tentatively scheduled for the first week of October. Once we have more information from them, we'll be able to come up with our project approach and begin working on it. Until that time we will continue to expand our knowledge base of the relevant libraries used by Principal as well as continue to study the scripts they provide to us.

Plans for Upcoming Reporting Period
Josiah Anderson
- Analyze sample scripts to find processes that could be automated or made more efficient

Doh Yun Kim
- Plan rough ideas/prototypes we can show based off of R and Python research
- More research into Python and R
- Revised interview questions to obtain better information based on current research

Gabriel Klein
- Continue to study Principal scripts
    - Look further into numpy, pandas, and sklearn
        - StandardScalar
        - LabelEncoder
        - mutual_info_classif
        - mutual_info_regression
        - numpy.corrcoef
        - make_scorer
    - Understand broad purpose of each section in rolling window script
        - splitting testing/training data
        - preventing contamination
        - engineering non-linear features

       - removing correlated columns
       - cross validation
       - feature selection
    - If time, begin studying portfolio simulation script

Drake Mossman
- High-level familiarization with R
   - Set up R development environment
   - Look at example R scripts from Principal
- Do initial planning for requirements and timeline

Jacob Richards
- Continue spending time on Python and R
- Go through data and examples provided by Principal

Nathan Schaffer:
- Research and experiment with R
- Research more about the example code provided by Principal
- Help enumerate tasks to report to Principal for weekly RASIC updates

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Josiah Anderson	R/Python research, client presentation, project problem/deliverables	8	15
Doh Yun Kim	Risk management, sklearn, parsing data	8	14.5
Gabriel Klein	Python research, numpy, pandas, sklearn, rolling window script	9	16
Drake Mossman	Pandas dataframes, sklearn, machine learning, Principal scripts	7.5	13.5
Nathan Schaffer	Python research, example scripts, project management	7	11
Jacob Richards	Python/R research, example scripts	5	9

Gitlab Activity Summary
Nothing to report.