MS-E2177 - Seminar on Case Studies in Operations Research

Applying Advanced Analytics in Asset Allocation

Interim Report

Client:
Varma

Team:
Leevi Olander (Project Manager)
Julius Lind
Samuli Mäkinen
Rasmus Heikkilä

April 23, 2019

1 Project Status and Accomplished Tasks

Our project was initially divided in 3 tasks:

- Exploratory data analysis and data processing
- Clustering of indicators & time periods and factor analysis
- Predictive modeling and benchmarking

Of these three tasks the first two have been completed, but further data processing will likely be needed as we begin implementing predictive models.

During exploratory data analysis all of the data was visually inspected to identify outliers or changes in data reporting (i.e changes in the scale of survey questions). No outliers or changes were found. Some time series were found that contained no data for some time periods, and were dropped from the dataset. Since the time-series were of differing lengths, some beginning as early as 1998 and some as late as 2011, we restricted the data set to only contain series that had data starting from 2005. This left us with a dataset consisting of 56 time series out of the original 60. Finally the time series were transformed to the same frequency by taking monthly averages from the daily and weekly series.

The clustering of indicators & time periods and factor analysis have provided valuable insight for the continuation of the project. For instance, the cross correlation has successfully identified lag based relationships, the principal component analysis has been used to decrease the number of input dimensions and the hierarchical clustering has identified which indicators are the most similar to each other.

Extensive literature reviews have been conducted regarding predictive models. Based on the literature reviews we are currently discussing some modeling choices with Varma, after which we can begin implementing the actual models. Additionally, the scope and objectives of the project were clearly defined by Varma at the start of the project. The workload and feasibility of the original objectives has been reasonable, so no changes have been made on that front.

2 Remainder of the Project & Schedule

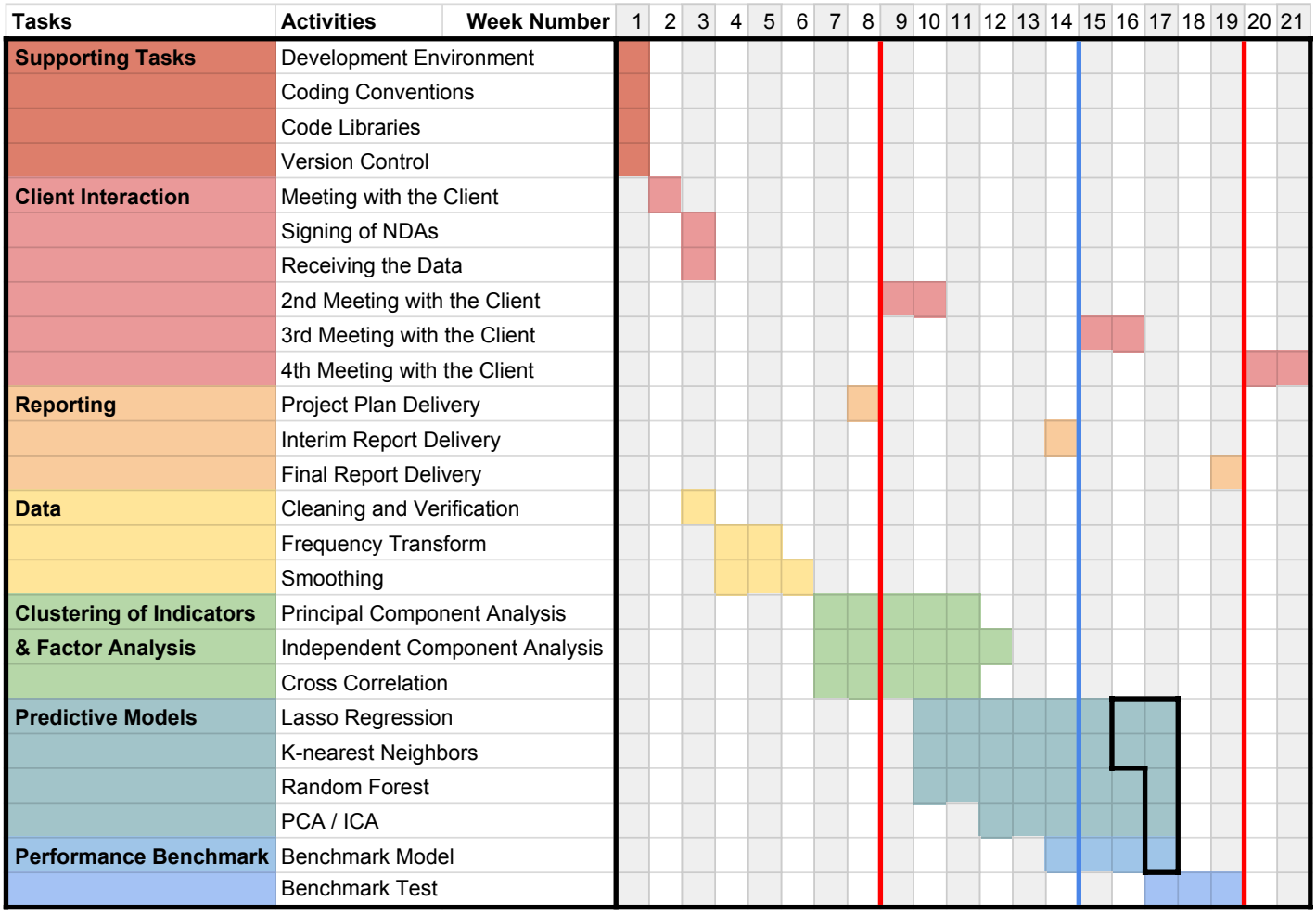

Figure 1: A Gantt chart of the revised project schedule, where the area surrounded by black borders has been updated. The week numbers corresponds to the weeks of the course MS-E2177, which started on 11.01.2019. The red vertical lines show the deadline for deliverables and the blue line symbolises the current deadline.

The project has progressed mostly according to the original schedule, as shown in 1. The first two tasks outlined in the previous section were completed by their planned deadlines. The schedule for predictive modeling has however changed. In our original schedule we had budgeted most of the time

into implementing different methods for predictive modeling such as lasso regression and K-nearest neighbors. However we quickly noticed that the actual implementation of the models with different methods takes very little time. The time consuming part has been defining the predictive models.

When defining our models several nontrivial choices have to be made. Firstly, how to use our data. Most regression methods assume that the input data is stationary, and the data is usually stationarized by taking the relative difference between subsequent points. This has the downside of losing information about the magnitude of the variable, which can be more important than the relative change for some variables. To define which variables to stationarize, we are heavily relying on Varma's expertise of financial markets.

As defining the predictive models has taken longer than expected, most of the remaining time has been budgeted towards model design. Implementing the models will be done along with model design until we obtain acceptable results.

3 Updated Risk Management Plan

Risk	Likelihood	Impact	Effect	Mitigation measures
Model fails in performance benchmark	Very High	High	Model is not useful to the client.	Testing multiple models and careful study of relevant literature.
Too large workload	Medium	Medium	Lower quality and delays in project schedule.	Start with a small number of models, expand the scope only if time permits. Maintain active discussion with the client about task prioritization.
Data quality issues	Low	Medium	Misleading, incorrect or inaccurate results.	Understanding the limitations both in data and methods used.
Team member absence / inactivity	Very Low	High	High workload for other team members.	Good communication between the project manager and the rest of the team. Clear schedule and fast reaction to delays.
Communication issues with the client	Very Low	Medium	Result is not what the client wanted.	Good communication with client by email and frequent meetings with the client.
Issues with computational resources	Very low	Low	Delays in project schedule.	Use algorithms with low enough computational burden.

Table 1: Updated risks related to the project. The background colors show whether the risk has either increased (red) or decreased (blue) from the previous assessment.

Table 1 shows our updated risk table. With the exception of model failure, the likelihood of all risks has decreased or stayed the same. The likelihood of model failure has increased from high to very high, as mostly poorly performing models were found during literature reviews. As the data has been thoroughly inspected with no issues emerging, it is unlikely that data quality would cause issues in the future. And since communication within our team and with Varma has been excellent, the these two risks were lowered.