Sho Tsubakiyama

MSc Business Analytics 2018 Class

Public Project Summary

Learning by Experience: Building a Decision Tree to QC and Resolve Data Issues

Table of Contents:

- I. Executive Summary
- II. Client background
- III. Issue addressed during project
- IV. Project timeline
- V. Conclusion

I. Executive Summary

My project involved the development and leadership of a data production enhancement at a large asset management firm. Besides the technical development of an SQL and UNIX based logic that eventually was programmed into an automated C script, the project was largely an exercise in leadership and taught me an immeasurable amount regarding what are the challenges of these kinds of projects. Also, it provided me with great exposure that will certainly lead to a positive review at the end of the year.

II. Client Background

The client for whom I conducted this project is a multinational asset management firm. The department in which I conducted this project is responsible for the delivery of various risk and performance analytics to portfolio managers and other stakeholders. For example, with this data, a portfolio manager can measure the risk of specific securities and asset classes or how their portfolio return compares against a certain benchmark. Therefore, the integrity of this data is extremely important since the management of large sums of money depends upon this data.

Almost every component of the data production process for these analytics is automated. Position are loaded, risk is calculated, and reports are sent to clients all by automated processes. Therefore, my client

team is not responsible for calculating analytics like exposure, duration, and returns manually. Instead, they are responsible for overseeing these automated processes, identifying and resolving data issues, and enhancing the data production processes.

III. Issue addressed during project

As addressed above, my client oversees an enormous and complex data production process. One part of this process is the quality check of the data being produced. Automated scripts check that changes are within a reasonable tolerance, did any processes fail, and is any output data missing. These issues are flagged to analysts to manually verify or resolve. For example, these issues could include alarming net asset value changes for a portfolio, suspicious changes in analytics like duration or value at risk, or simply missing analytics due to production or set up issues.

As such, my project focused on a specific issue by which trade returns failed to be calculated during production. My project was part of a larger initiative by which my client was building decision trees that would mimic analyst logic and resolve or at least provide relevant information to analysts to minimize the amount of manual labor that they would expend addressing the issue. Rather than the technical aspects of this project, my capstone project focuses on what the process of implementing this data solution was like from start to finish within the context of a large data production process and company with over 10,000 employees.

IV. Project Timeline

Month 1: Development of the Tree

This stage involved a lot of working alone to understand the specific data issue of why this trade return was not being calculated. If anything, this was the easiest stage of the project as I could work largely alone and the basic logic was relatively easy construct. Primarily, I used SQL logic to query different databases to make checks that would mimic the analysis of data analysts who normally investigate the issue. However, as I later found out, there was a substantial volume of outlier cases that did not respect the straightforward logic.

Months 2-3: Building the Decision Tree

Once I finalized the tree with a specialist in the topic and completed the SQL queries that each node would use, I worked with a small team of developers to build the SQL logic into a C-based program that would automatically run as a part of the standard production process. The major challenge of this process was communication and working with the developer release systems of the company. As I don't know C, I had to rely upon the team and could do little to resolve issues when the tree did not work as expected.

Months 4-6: Testing, Implementing, and Refining the Tree

Already in the above "building stage," we were conducting many tests. However, this final stage saw more testing across the company and the further development of the tree to not only analyze issues but to also automatically resolve them via Unix based commands. This significantly increased the difficulty of the project as this forced me to recognize that there were many outlier cases that were not being captured by the tree. These outlier issues were considerably more difficult than the rest of the cases, and I am still investigating them. Also, communicating to the rest of the team why the tree did not work as anticipated in these cases was very difficult. Regardless, the tree has now been implemented in its first stage and resolving 60-80% of these issues. Itis planned to be refined further over the next months.

V. Conclusion and Lessons Learned

In conclusion, this project tested me considerably and taught me a lot about leading a data production enhancement project in a large company with many stakeholders. Summarized into a few points, here are some of the major lessons learned:

- **Don't expect things to work perfectly from the start (or ever).** This expectation management is important for yourself and stakeholders.
- **Relationships are key**. Know your allies and who is responsible to see that your project is executed. Also, don't take people for granted and make sure that you nurture relationships (for the sake of your current project and for those in the future).
- Appreciate the career value of these projects outside your day-to-day role. Due to this project. I have learned a lot about a specific topic that not many people have expertise in in the company. Also, I have met and built relationships with many people who I would have not otherwise known. Finally, I have received a lot of positive attention around the company.